

89966

STIC-Biotech/ChemLib

Fr m: Chan, Christina
Sent: Wednesday, March 26, 2003 11:08 AM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: In Re:09884465

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Ford, Vanessa
Sent: Wednesday, March 26, 2003 10:57 AM
To: Chan, Christina
Subject: In Re:09884465

Please search SEQ ID NO:332. Please include interference searches. Please rush.

Vanessa L. Ford
Biotechnology Patent Examiner
Office: CM1 8A16
Mailbox: CM1 8E12
Phone: 703.308.4735

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/27
Date Completed: 3/28
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:17 ; Search time 49 Seconds
(without alignments)
1777.506 Million cell updates/sec

Title: US-09-884-465a-332
Perfect score: 4728
Sequence: 1 MQIYTDDEIQVAKLAGRYT.....IELRLPSGEVKKNLSDFFIA 906
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3357	71.0	1039	2	H95115	conserved hypotet
2	3356	71.0	1039	2	D97985	hypothetical prote
3	1307.5	27.7	839	2	G95115	conserved hypotet
4	1245	26.3	853	2	C97985	hypothetical prote
5	946.5	20.0	819	2	B95136	conserved domain p
6	946	20.0	855	2	D98004	histidine Motif-Co
7	636	13.5	802	2	C95136	conserved domain p
8	636	13.5	828	2	E98004	hypothetical prote
9	223.5	4.7	822	2	T46758	hypothetical 92.4k
10	202.5	4.3	2269	2	T86777	rhodry protein -
11	201.5	4.3	1873	2	T30944	surface protein pr
12	201	4.3	1790	2	S67593	transport protein
13	201	4.3	2004	2	F95133	immunoglobulin A1
14	197	4.2	1478	2	S20117	protein kinase BCK
15	197	4.2	3488	2	T34418	hypothetical prote
16	194.5	4.1	1134	2	A60234	Iga Fc receptor pr
17	194.5	4.1	1164	1	FCOAG	Iga Fc receptor pr
18	193.5	4.1	1125	2	E90598	membrane nuclease,
19	192	4.1	1871	2	D96796	probable heat shoc
20	190	4.0	4688	2	F92885	hypothetical prote
21	189.5	4.0	2748	2	S57976	nuclear migration
22	189.5	4.0	5005	2	F82884	hypothetical prote
23	189.5	4.0	5327	2	T13564	microtubule-associ
24	186	3.9	1640	2	A24594	probable major sur
25	186	3.9	1658	2	S55101	hypothetical prote
26	185.5	3.9	1983	2	G96643	hypothetical prote
27	185.5	3.9	2109	2	E99066	protein H05009.1
28	185.5	3.9	2109	2	T33247	hypothetical prote
29	185	3.9	2485	1	H71621	serine/threonine-s

30	184.5	3.9	1963	2	B98002	Iga-specific metal
31	183	3.9	2218	2	B84683	hypothetical prote
32	182	3.8	2450	2	S71625	protein-tyrosine-p
33	181.5	3.8	1139	1	E64234	cytadherence-acces
34	181	3.8	1939	2	T18372	repeat organellar
35	181	3.8	1979	2	JW0059	mtprd protein - mo
36	181	3.8	2481	2	D90011	FmtB protein [impo
37	180.5	3.8	1185	2	A42404	collagen adhesin -
38	180	3.8	1371	2	A45555	glutamate rich pro
39	180	3.8	1639	2	S05603	major merzozolite au
40	179.5	3.8	2447	2	T18670	hypothetical prote
41	179	3.8	1969	2	T38495	hypothetical prote
42	178.5	3.8	763	2	T08929	hypothetical prote
43	177.5	3.8	1881	2	H95076	zinc metalloprotel
44	177	3.7	1305	2	H41662	150K mating aggreg
45	176.5	3.7	2195	2	S61103	SEC16 protein - ye

ALIGNMENTS

RESULT 1

H95115
conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Retcelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequences of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1004

Query Match 71.0%; Score 3357; DB 2; Length 1039;
Best Local Similarity 72.3%; Pred. No. 2,1e-157;
Matches 687; Conservative 40; Mismatches 95; Indels 128; Gaps 12;

QY	7	DDEIQVAKLAGRYTTEDGYIF	-----DTS-----	WIKKDSLSERAAAAQA	47
DB	168	NSNVAVARSGRYTTNDGVFNPA	DIETGNAYIVPHGHYHYIPKSDLSAS	ELAAAKA	237
QY	48	YAKERGLTP	-----PSTDHQDSGNTAKGA	-----EATYRVKAAKVP	87
DB	228	HLAGKNQPSQLSYSTASDNDNTQ	SVAKGTSKPANKSENLOSLLKELYS	PSAQRYSES	287
QY	88	DRMPYNLQTVTKNGSLIPSDV	HYHNKFEWFDGLYEAPRGYSLD	LATVKYVPE	147
DB	288	DGLVFDPAKIIISPTNGVAIPH	GDHYHFIPI	-----KLSALEEKIANN	---VPI 334
QY	148	RNASDHVRKNAQDQSKPDEK	HEDEVSEPTPESEKENHAGLNPS	ADNLKPKSTDTEE	207
DB	335	SGTGSTVSN	-----AKPNEV	-----SSLGSLSSN	---PSSLTTS 367
QY	208	TEEEAEDTTDEAIEPCTPSI	RONMETLGLKSSLLLGKDNNT	ISAEDVSLALAKESQ	267
DB	368	KE	-----LSSASDGYIFNPKD	-----IVEETA	389
QY	268	RAPI	-----OGPOIGOPTLPNNSLAT	PPSLPINCPTSHEKHEEDG	YGFEDANI 316
DB	390	TAIYVHGDFHYIPKSNQIG	OPTLPNNSLATPPSLPINCPTS	HEKHEEDG	YGFEDANI 449
QY	317	IAEDSEGFVMSHGDSNHVFF	FKDKLTDEEQKAAQKHLEEVK	TSNGLDLSLSSEHQDYP	PGNA 376

Db 450 IADESGFVMSHGHNHYFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSEHQDYPNSA 509
 QY 377 KEMKLDKKIEEKIAGIKOYGVKRESIVNKEKNAIIPSGDHHADPIDEHKPVGIGH 436
 Db 510 KEMKLDKKIEEKIAGIKOYGVKRESIVNKEKNAIIPSGDHHADPIDEHKPVGIGH 569
 QY 437 SHSNYELFPKPEGVAKKGNKYTTGEELTNVNNLKNSTFNNQFTLANGOKRVYSFSP 496
 Db 570 SHSNYELFPKPEGVAKKGNKYTTGEELTNVNNLKNSTFNNQFTLANGOKRVYSFSP 629
 QY 497 ELEKKGILNMLVKLITPDGKVLKSGVFGVGVGNIANFELDPYLPQGTFKYTIASKD 556
 Db 630 ELEKKGILNMLVKLITPDGKVLKSGVFGVGVGNIANFELDPYLPQGTFKYTIASKD 689
 QY 557 YPEVSDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 616
 Db 690 YPEVSDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 749
 QY 617 NAYLENNYKVGEEKILPILNMGTTTGNKIPVTFMANAYLDNOSTYIVVEPILKRENO 676
 Db 750 NAYLENNYKVGEEKILPILNMGTTTGNKIPVTFMANAYLDNOSTYIVVEPILKRENO 809
 QY 677 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDVP 736
 Db 810 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDVP 869
 QY 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 796
 Db 870 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 929
 QY 797 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGVNGSDPMLDPALEAP 856
 Db 930 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGVNGSDPMLDPALEAP 989
 QY 857 AVDPVQEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDFIA 906
 Db 990 AVDPVQEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDLIA 1039
 RESULT 2
 D97985
 hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: D97985
 R:Roskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 y, P.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: D97985
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1039 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:gl5458515; GSPDB:GN00174
 A:Genetics:
 A:Gene: phtE
 Query Match 71.0%; Score 3356; DB 2; Length 1039;
 Best Local Similarity 72.0%; Pred. No. 2.4e-157;
 Matches 684; Conservative 43; Mismatches 95; Indels 128; Gaps 11;
 QY 7 DDEIQVAKLAGKTTEDGYTF-----DTS-----WIKKDSISEARRAAQA 47
 Db 168 NSNVAVARSOGRTTNDGVFNADIEDTGNAYIVPHGHYHYIPKPSLSASLAAKA 227
 QY 48 YAKEKGLTP-----PSTDHQDSGNTAKGA-----BAIYNKVAKAKVPL 87
 Db 228 HLAGKNQPSQLSYSTASDNNQTSVAKGSTKSPANKSENQSLKELDPSAQRYSES 287
 QY 88 DRPYNLQYVEVKNGLIIPSDHYHNKFEWDEGLYAPKGYSLDILATYKYTYEP 147

Db 288 DGLVDPKAKIISRTPNGVAIIPHGDHYHFIPYS-----KLSALEKIA----- 329
 QY 148 RNASDRVKNKAKQDQSKPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSDTTEE 207
 Db 330 RRVPISTGTVSTNAKPNEV-----SSLSLSN-----PSSLTYS 367
 QY 208 TEEAEEDTDEAIEPOTPSIRQNAMETLGLKSSLLGTNDNNTISAEVDSLALAKESQ 267
 Db 368 KE-----LSSASDGYIFNPKD-----IVEETA 389
 QY 268 PAPI-----OGPOIGOPTLPNNSLATPSLPINPGTSHKHEEDSYGFDANRI 316
 Db 390 TAYIVRHGDHFHYIPKSNQIGQFTLPNNSLATPSLPINPGTSHKHEEDSYGFDANRI 449
 QY 317 IADESGFVMSHGSDSNHYFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSEHQDYPNSA 376
 Db 450 IADESGFVMSHGSDSNHYFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSEHQDYPNSA 509
 QY 377 KEMKLDKKIEEKIAGIKOYGVKRESIVNKEKNAIIPSGDHHADPIDEHKPVGIGH 436
 Db 510 KEMKLDKKIEEKIAGIKOYGVKRESIVNKEKNAIIPSGDHHADPIDEHKPVGIGH 569
 QY 437 SHSNYELFPKPEGVAKKGNKYTTGEELTNVNNLKNSTFNNQFTLANGOKRVYSFSP 496
 Db 570 SHSNYELFPKPEGVAKKGNKYTTGEELTNVNNLKNSTFNNQFTLANGOKRVYSFSP 629
 QY 497 ELEKKGILNMLVKLITPDGKVLKSGVFGVGVGNIANFELDPYLPQGTFKYTIASKD 556
 Db 630 ELEKKGILNMLVKLITPDGKVLKSGVFGVGVGNIANFELDPYLPQGTFKYTIASKD 689
 QY 557 YPEVSDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 616
 Db 690 YPEVSDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 749
 QY 617 NAYLENNYKVGEEKILPILNMGTTTGNKIPVTFMANAYLDNOSTYIVVEPILKRENO 676
 Db 750 NAYLENNYKVGEEKILPILNMGTTTGNKIPVTFMANAYLDNOSTYIVVEPILKRENO 809
 QY 677 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDVP 736
 Db 810 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDVP 869
 QY 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 796
 Db 870 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 929
 QY 797 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGVNGSDPMLDPALEAP 856
 Db 930 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGVNGSDPMLDPALEAP 989
 QY 857 AVDPVQEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDFIA 906
 Db 990 AVDPVQEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDLIA 1039
 RESULT 3
 G95115
 conserved hypothetical protein SP1003 [imported] - Streptococcus pneumoniae (strain T
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: G95115
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
 son, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
 A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: G95115
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <KUR>

[illegible]

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match	13.5%	Score	636;	DB	2;	Length	828;
Best Local Similarity	45.3%	Pred.	No. 9.3e-24;				
Matches	140;	Conservative	40;	Mismatches	75;	Indels	54;
Gaps							

Qy	2	QITVDDIEIQVAKLGYTTEDGYFD-----	-----TSWIKKDSLSAEAR	42
Db	539	QIETVEDVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLSDEK	598	
Qy	43	AAQAQYAKKGLTPPSITDHQDSGTEAKGAEAIIYNRVKAAKKVPYLDRMPNYLQYTVVKN	1020	
Db	599	VAQAQYTKKGIPLPSDPAQYKAPNTGDSAAAIYNRVKEKRIPLVRLPYVHEHTVEKN	658	
Qy	103	GLSLIPSDYDHVHNKTFEWFDEGLYAKPKGSLEDLLATVKYVE-----	-----PR-----NAS	151
Db	659	GNLIIPKDHVHNKTFWFDDHTYKAAPNGVTLEDLPATIKYIVHEPDERPHSDNGWGNAS	718	
Qy	152	DHVRKNKAQDSKPEDKEHDEVSEPTHPSEDEKENHAGLNPSADNLKYPSTDTETEETEE	2111	
Db	719	EHV-LGKKDHSDPNKFNKADE--EPVEETPAEPE-----	-----VPQVETKVEAQ	762

Db	763	LK----	EAEVLAKVTDSS	LFKANATET	LAGLRNLT	LQINDNNS	IMAEAEKL	ALLKGSN	818
Qy	268	PAPIQQPQI	276						
Db	819	PSSVSKKI	827						

Db 399 OSFNDKSLNETKNSIEKYO--NINTLKVKVDEYIKVKCKTESITKSFSSQOTILKMDLN 457
 QY 392 -----GIMKOYGVKRESIVYKKE---KNALIPSGDHHADP-----IDHKP 431
 Db 458 ONIKVKTNSIDKSYIEKFFQIITGKOTKLENAFTFSLNHHANNELIKYSDUKAN 517
 QY 432 VGIGHSHSNYELPKEGVAK--REGKRVYTGEBLTNNVLLKRNSTNNFNTILANGKR 489
 Db 518 LGINEENMLYNQFTEKRTFNDEKKN--IHINEISIKIEIKHAIYN-----564
 QY 490 VSFSPPPLEKLGIMLVKLLITDGVKLVKSGVKGEGVGNVTANFELQYLPQGTFK 549
 Db 565 -----ISETEIEIGIN-----IESLNTKVFKEKVENYTNL-----NRK 599
 QY 550 YTIASKDYFVSVDGTFVTPTSLAYKMASQTIFFPFHAGDTYLRVNPQ--FAVPGTDAV 608
 Db 600 EKLKHYDFSGKEG-----NIKYT-----DKIKINDIMAYSQIDQHI 640
 QY 609 RVDFPHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIYEV 668
 Db 641 NGLDDIQKKS---ESY-VSEMEQINKLEK-----VSNTESISDN-----V 677
 QY 669 PILKEKQ-----TDKPSILPQFKNRAQENSKL---DERVEEPTKS-EKVEKELSETG 719
 Db 678 EGIRKKQIIVTKDK-----KNYEEINKLLSEIKEDNTSLEKVKDINLS-YG 729
 QY 720 NSTNSLLEEVYTPDVPQEVAKFAESYGMKLENVLFMDGTIELYLPSGEVIKKNADF 779
 Db 730 QNLGNLFLEQIDEEKKKAENTIKSMEAYIDDLONIK-KKQSEIETEMDIRMDINKMEAL 788
 QY 780 TGEAPQNGENKPSNGKSVGTGVENOPT-----ENKPADSLPAPNEKP 824
 Db 789 KISHDDKKCHKSKNKHENISDIYKSKIIQDSRESINDIKKLNQKVSQSN---845
 QY 825 VKPNSDNGMLNPEGNGVSDPMDLPALEAPAVDPQVEKLEKFTASYGLGL---DSVIF 881
 Db 846 ---NSINDOCLNVAIYNILKALKIKKIIDKVKETSEIEKKNKINDELNNSKVIK 902
 QY 882 NMOGTIELR-----LPSGEVIKKNL 901
 Db 903 KIEDLSLKECRSKINSTLDDKDIDECIKINVLKKN 940
 RESULT 11
 T30944
 surface protein precursor - Enterococcus faecalis
 C:Species: Enterococcus faecalis
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30944
 R/Shankar, V.; Baghdadyan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
 Infect. Immun. 67, 193-200, 1999
 A>Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene end
 A:Reference number: Z20943; MUID:99081742; PMID:9864215
 A:Accession: T30944
 A>Status: preliminary; translated from GB/EMBL/DDB
 A:Molecule type: DNA
 A:Residues: 1-1873 <SHA>
 A:Cross-references: EMBL:AF034779; NID:G3873186; PID:G3873187; PIDN:AAD09858.1
 Query Match 4.3%; Score 201.5; DB 2; Length 1873;
 Best Local Similarity 21.8%; Pred. No. 0.071;
 Matches 209; Conservative 116; Mismatches 337; Indels 297; Gaps 53;
 QY 2 QITYTDDIOVAKLAGYTT-EDGYIFDTSWIKKDSLSAEARAAQAQAKELGTPPSTD 60
 Db 842 KVTF-----AGBGTIESG---TTFVAVKDGVSFLPDLKPLVKAD-----880
 QY 61 HQDSNGTNEAKGAEIYNVKA-----AKVPLDRPYNLYQYTVKNGSLIIPSDHYH 114
 Db 881 -----GYTDKWPPEATOPIKADDTFVSSATKLDI-----IENPGDNIPA--GYH 925
 QY 115 NIKFENFDEG-----LYEAPKGYSL-EDLLATVKY---YVEPRNASHVRKNKADQ- 161

Db 926 KVTF-AGEBCTSIESGTTTFAVKGVSFLPEDKLPVLKAKDGYTDKWPGEATOPIKADDT 984
 QY 162 -----DSKPDDEK---EHDEVSEPTHPESDEKE---NHAGLNPSADNLKPSDTE 206
 Db 985 EFVSSATKLDKSDAKRYTPEGOKVVTTELKNEPDASEGINKKDLPRDAKYTKWKYDIS 1044
 QY 207 ETBEE-----AEDTDEAEIPGTPSIRQNAME-----TLTGLKSLLAGTDDNNTISA 254
 Db 1045 TAGNKGTVVTVYSDSSDEVDVTVTDNRSDADKYEPVEGEK--VEIGGVDTL---1099
 QY 255 EVSLALLAKESQAPIQGPIQOPLPNNSLATPSPSLPINFPTSHEKEEDGYGFAN 314
 Db 1100 --DNVTNL-----PTLPQGTIVT-----DVTGGTIDTNPQNY-----1131
 QY 315 RIIAEDSGFVMSHGSHNYFFFKDLTEQIKA-----AQKH---LEEVTSHNGLDL 365
 Db 1132 -----EGVETVTPGT-----KDTVKVPVEVTDNRSDADKYEPVEGEKVEIGKVDL 1180
 QY 366 SSheQDYP--GNAKEMKD-----LQKIEKIAGIKQYGVKRESIVVKNKNAIYPS 417
 Db 1181 TDNVTNLPTLPQGTIVTDTVTPGCTIDTNPQNYEGVIE-----VTPD 1223
 QY 418 GDHHADPIDEHK-PVGIGHSHSNYELFKPE-EGVAKKEGKNYVTGEELNNVLLKNST 475
 Db 1224 GTR-----DTVKVPVEVTDNRSDADKYEPVEGEKVEIGKVDLTDNVTNLPTLPQGT 1277
 QY 476 FNN-----QNTFLANGQKRVSFSPPELEKLGINMLVKL---ITPDGKVLKESGKV 535
 Db 1278 VDTVTPGCTIDTNPQNYEGVIEVTPD-----GTRDVKVPVEVTDNRSDADKYTPMV 1331
 QY 526 FSGEV-----GNIAFELDQYLPQGTFFKTYTISKDYPVSVGDTFTVTPSLAYKM 576
 Db 1332 EGKVEIGKGVLDTDNVTNL---PTLPQGT---TVT-----DVTGGTIDTNPQNYEG 1379
 QY 577 ASOTIFPFHAGDTYLRVNPQFAVPGTDAVRFDFHGNAYLENNY-----KV---526
 Db 1380 VIEV-----TY-----PDGTDVKVPVEVTDNRSDADKYEPVEGEKVEIG 1421
 QY 627 GEIKLP-----IPKLNQGT---RTAGNKIPVTFMANAYLDNQSY-----IVEVPILE 672
 Db 1422 GKVDLTDNVTNLPTLPQGTIVTDTVTPGCTIDTNPQNYEGVIEVTPDGTVDKTVKPV 1481
 QY 673 KENQTDKPSILPOFKRKAQENSKLDEKVEEPTSEKVE-KEKLSGTGNSSTNLEEVP 731
 Db 1482 TDNRSDA-----DKVEPTVEGEKVEIGKVDLTDNVTNLPTLPQGT 1522
 QY 732 TVDPVQEKVAKFAESYGMKLENVLFMDGTIELYLPSSG--EVK-----KNADTGEA 793
 Db 1523 TITDVTGP-----GTIDTNPQNYEGVIEVTPDGTVDKTVKVPVEVTDNRSDADKYE 1574
 QY 784 PQNGENKPSNGKVS-TGTVENQPTENKPADSLPAPNEKPKVPKPNSTNDGMLNPEGN 841
 Db 1575 PVEGE-KVEIGKVDLTDNVTNLPT-----LPGGTITDVTGGTIDT---NTPGN 1622
 RESULT 12
 S67593
 transport protein US01 - Yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein D2552; protein YDL058w
 C:Species: Saccharomyces cerevisiae
 C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
 C:Accession: S67593; A38455; S30782
 R:Blöcker, H.; Brandt, P.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67587
 A:Accession: S67593
 A:Molecule type: DNA
 A:Residues: 1-1790 <BL0>
 A:Cross-references: EMBL:Z74106; NID:G1431058; PID:e253003; PID:G1431059; MIPS:YDL058
 A:Experimental source: strain S288C
 R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
 J. Cell Biol. 113, 245-260, 1991
 A>Title: A cytoskeleton-related gene, US01, is required for intracellular protein tra

A:Reference number: A38455; MUID:91185402; PMID:2010462

A:Accession: A38455

A:Molecule type: DNA

A:Residues: 1-389, TA, 392-724, S', 726-1790 <NA>

A:Cross-references: GB:X54378; MID:94777; PID:CAA38253.1; PID:94778

A>Note: the authors translated the codon ACT for residue 768 as Ile

R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.

submitted to the EMBL Data Library, February 1993

A:Description: An integrin analogue in Saccharomyces cerevisiae.

A:Reference number: S30782

A:Accession: S30782

A:Molecule type: DNA

A:Residues: 71-846, E', 848-923, K', 925-1252, I', 1254-1318, V', 1320-1460, S', 1462-1580, S

A:Cross-references: EMBL:L03188

C:Genetics:

A:Gene: SGD:US01; INT1

A:Cross-references: SGD:S0002216; MIPS:YDI058W

A:Map position: 4L

C:Keywords: coiled coil; transmembrane protein

F:326-342/Domain: transmembrane #status predicted <TM1>

F:394-410/Domain: transmembrane #status predicted <TM2>

F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 4.3%; Score 201; DB 2; Length 1790;

Best Local Similarity 19.3%; Pred. No. 0.071;

Matches 183; Conservative 132; Mismatches 328; Indels 304; Gaps 40;

QY 67 TEAKGAEALYNVKAAKVPLDPMYINLQYTVVKNGLIIPSYD-----HYHNKFE 119

DB 650 TTKLQDNVASHIKQFK-----DSYFSKYDMNEDSILTPELDTGLPKRYVFTYFIQ 702

QY 120 WFDEGLYAPKGYSLLEDLLATVKYVPEPNASDHVKKNADQSDPKDEKHEV----- 174

DB 703 LFNENIYRI-----RTALSH-----DPEEPINKISVEEVEKLOR 737

QY 175 -----SEPHPSDEK-----ENHAGLNPSADNLKPKSTDTDE 207

DB 738 QCTKLKGEITSLOTETESTHENLTKLIATNEHKELDEKYQILNSSHSL-KENFSILE 796

QY 208 TE-EAEDTTDAEIPGTPSIRQNAMETITGLKSSLLGTCKNNITSAEVDLSLLALKES 266

DB 797 TELKNVRDSLDE-----MTQLRDVLTKDKENQTALEYKSTI-----HK 836

QY 267 QPAPIQGQPGIQTPLNPSLATPSPSLPINPGTSHKHEEDG-----YGFDA NRILA 318

DB 837 QEDSIKLEKLETI-----LSQKKAEDGINKMGKDLFALSREMQAV 879

QY 319 EDESGFVMSHGDS---NHVFFKDLTEE-----QTKAAQKHLEVKVTHSHGL---DSL 365

DB 880 EENCNLOKREKDSNVHOKETKSLKEDIAAKITEIKAINLENLEEMKIQCNLSKEKEHI 939

QY 366 SSEDYDPCGNAXEMKDLKKIEKTAGINKQY---GVKRESIV---VNKEKN-AIIYPSGD 419

DB 940 SHELVEYKSRFOSHNLVAKLLEKLSLANNYKDMQAEENESLIKAVESKNESLSQLSNL 999

QY 420 HHADPIDEHK---PVGIGHSHSNYLEF-----PEEGVAKKEGKNVYTGELTNV 467

DB 1000 ONKIDSMQEKENFOIERSIEKNIETQLAKTISLDTQETKEEIIISKSDSK-----DEYESQ 1055

QY 468 VNLLN-----STFNQNPETLANGOKRVSFSPPELE--KKIGINMLVKLIITPDGKVLBK 520

DB 1056 ISLLEKLEKLTATTANDENVNKISLTKTKEELEAEALAAAYKNLKNLETKLETSE-KALKE 1114

QY 521 V-----SGKYVGEVGVGNIANFELDQPLPGQTFKY--TIASKDP 558

DB 1115 VNEBEHLKEEKIQLEKATETATQQLNSLRANLESLEKEHEHDLAQLKKYEQIOANK---- 1171

QY 559 EYSYDCTFTVPTSLAYKMASQIFYPFHAGDTLYLRVNPQF-AVPKGTALVRVDFEFGN 617

DB 1172 ERQYN-----EISQLNDEITSTQENESIKKKNLEGE 1206

QY 618 AYLENNYKVGKILPKPLNQCTTTAGNKIPVTPMANAYLDNQSTIYVEVPILEKNOT 677

DB 1207 V-----KAMKSTSEOSN-----LKKSEIDALNLOIKE---LKKNET 1241

QY 678 DRPSILPOFK-----RNKAQNSKLDEKVE--EPTSKVKEKLSE--- 717

DB 1242 NEASLESIKSVESETVKIKELQDCNFKREKEVSELEDKLASEDKNSKYLELOKESKI 1301

QY 718 ----TGNSTNSTLEEVPTDPOVEK-----VAKPAESYGMKLENVLFNM 758

DB 1302 KEELDAKTKELIKQLEKITNLKAKESSELSRLKKTSSERKNAEPOLEKLN----- 1356

QY 759 DGTIELYLPFSGEVIKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPE 818

DB 1357 -----EQIKNQK-----PEKERKLLNEGS-STITQEYSKINTLEDELIR 1396

QY 819 APNEKPVTPENSTDNGLMNPENGVSDPMDLPALAEAPAVDPVQEKL 865

DB 1397 LQNELKAKEDINTRSELEKSVLSNDELLE---EKQNTIKSLQDEI 1440

RESULT 13

F95133

Immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001

C:Accession: F95133

R:Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H

non, T.D.; Mayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95133

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-2004 <CUR>

A:Cross-references: GB:AEO05672; PIDN:AAK75263.1; PID:g14972632; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1154

C:Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase

Query Match 4.3%; Score 201; DB 2; Length 2004;

Best Local Similarity 21.0%; Pred. No. 0.083;

Matches 185; Conservative 112; Mismatches 250; Indels 332; Gaps 48;

QY 156 KKAQODSPDEKDE--HDEVSEPHPESEDEKENHAGLNPSADNLKPKSTDTDETEEEAE 213

DB 51 KYVADSELSESEKKQLVYD---IPTVNDDETYTL-----VYKLSNQ----- 92

QY 214 DTTDAEIPGTPS-----IRQNAMETLGLKSSLLGTCKDNT 251

DB 93 ----LAELPNTGSKNERQALVAGASLAANGILIFAVSKKKVKNKTVLHLVLVAGINGVL 148

QY 252 ISAEVDSLALAK-----ESQAP-----IQGPQIGQPTLPN--NSLA 287

DB 149 VSVHALENHLLNNTDYELTSGEKLPLPKEISGYTYIGYKEGKTSSEVSQKSSVA 208

QY 288 TSPSLPINTPGTSHKHEEDGYGFDA-----RIADESGFVMSHGSHSNHYFFKDKIT 341

DB 209 TPT-----KQKVDYNNVTPNFVDHPSTVQAIOEQTPVSGSTKPTQVVEKPPS 256

QY 342 EQIKAAQKHLEVKVTHSHGLDSLSEHQDYDPGNAKEMKDLKKIEKTAGINKQYGVTR 401

DB 257 TELINPR-----KEEKQSDSQPOLAEH-----KNLETKKEEKIS----- 291

QY 402 ESIVYVKNKNAIIPSGDHHADPIDE-----HKPVGIGHSHSNHYLPKPEGVAKKEG 455

DB 292 -----PKRTGV-----NTLNPDQVLSGQINP-----ELLYRETM----- 325

QY 456 NKVYTGEEELTNVNLKNSITFNQNTLANGOKRVSFSPPELEKLGINN-LVKLIT-- 512

DB 326 TKIDFQEEI-----QENPOLAEGTVRV-----KQEGKLGKVKVIVRIFS 366

QY 513 -----PDGKVLKESGK--VFGE-----GV-----GNIANFELDQPLP 544
Db 367 KEESREIVSTTAPSPRIVEKGTGKTQVKEQPEGVHEDKQVSGAIVEPAI-QPELP 425
QY 545 GQTKFYIASKDYDEYDGTFTVPTSLAYKMASQTIFYFPACDQTYLRVNPQ-----FAV 600
Db 426 -----EAVVSDKGEVQD-----TLPEAVV-----TDKGET--EVQPESPDTWS 464
QY 601 PKGTDALVRVDEPHGNAYLENNKVEIK--LPPLK--NOGTRTAGNKIPVTFMANAY 657
Db 465 DKGPEQVAPLPEKGN-----IQVPEVETKTKGQPKET--EEVPV-----507
QY 658 LDNOSTVIVEPILEKENQTDKPSILPQKRN-----KAQENSKLDEYEEPTSEK-----709
Db 508 -----KPEETFPVNEGTTGTSI--QEAENPVQPAEESTNSEKV--SPDTSKNTGEV 559
QY 710 -----VEKEKLSGTGNTSNTSILEVPTVDPVQEVAKFAESVGMKLENVLF 756
Db 560 SSNPDSSTTSVGEKNKPEHNDKNSKNTSEKTYEVP--VNP-----597
QY 757 NMDGTIELYPSGEVIRKKNMADFGAPOGNGENKPSGKGV---STGTVENOPTENKP- 812
Db 598 -NEGIVE-----GTSNQETEPVQPAETQTSNGKIANENTGEVSKKPSDSKPP 645
QY 813 -ADSLPEAPNEKPKVPKENS-----TDNGHLPNE-----GNVSDPMLDPALEAPAYDPVQEK 864
Db 646 VEESNQPEKNGTATKPSNGTNGTENGQTEPEPSNGSTEDVSTESNTSNGNEEIKQE 705
QY 865 LEKTASVGLDSDVIFPMQDTIELRLPSGEVIRKKNLSD 903
Db 706 NE-----LDPKKVEPEKTELEP-----NVSD 728

RESULT 14
S20117
Protein kinase BCK1 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J0906; protein kinase SUK1; protein kinase SSP31; protein YH
C:Species: Saccharomyces cerevisiae
C:Date: 23-Apr-1993 #sequence, Revision 23-Apr-1993 #text, change 24-Sep-1999
C:Accession: S20117; S20298; S22285; S19061; JQ1432; S56872; S30794; JQ1118
R:Coogan, C.; Gehring, S.; Snyder, M.
Mol. Cell. Biol. 12, 1162-1178, 1992
A:Title: A synthetic lethal screen identifies SUK1, a novel protein kinase homolog impl
A:Reference number: S20117; MUID:92186847; PMID:1545797
A:Accession: S20117
A:Molecule type: DNA
A:Residues: 1-1478 <COS>
A:Cross-references: EMBL:M84389
A:Experimental source: strain S288C
R:Miosga, T.; Boles, E.; Schaff-Gerstenschlaeger, I.; Schmitt, S.; Zimmermann, F.K.
Yeast 10, 1481-1488, 1994
A:Title: Sequence and function analysis of a 9.74 kb fragment of Saccharomyces cerevisia
A:Reference number: S20298; MUID:95176706; PMID:7871887
A:Accession: S20298
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1478 <MO>
A:Cross-references: EMBL:X77923; MUID:9640004; PIDN:CAA54896.1; PID:9640009
R:Lee, K.S.; Levin, D.E.
Mol. Cell. Biol. 12, 172-182, 1992
A:Title: Dominant mutations in a gene encoding a putative protein kinase (BCK1) bypass t
A:Reference number: S22285; MUID:92107166; PMID:1729597
A:Accession: S22285
A:Molecule type: DNA
A:Residues: 1-58, 'I', 60-1478 <LEE>
A:Cross-references: EMBL:X60227
A:Experimental source: strain EG123
R:Lee, K.S.; Levin, D.E.
submitted to the EMBL Data Library, June 1991
A:Description: An extragenic suppressor of mutations in the S. cerevisiae protein kinase
A:Reference number: S19061
A:Accession: S19061

A:Molecule type: DNA
A:Residues: 1-58, 'I', 60-263, 'P', 265-278, 'I', 280-702, 'S', 707-708, 'KP', 714, 'VTMT', 715
A:Cross-references: EMBL:X60227; MUID:93414; PIDN:CAA42788.1; PID:93415
A:Experimental source: strain EG123
R:Irie, K.; Araki, H.; Oshima, Y.
Gene 108, 139-144, 1991
A:Title: A new protein kinase, SSP31, modulating the SMP3 gene-product involved in pl
A:Reference number: JQ1432; MUID:92104496; PMID:1840547
A:Accession: JQ1432
A:Molecule type: DNA
A:Residues: 149-1478 <IRI>
A:Cross-references: EMBL:D10389; DBJ:D90446
R:Miosga, T.; Schaff-Gerstenschlaeger, I.; Baur, A.; Boles, E.; Chalwatzis, N.; Four
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56855
A:Accession: S56872
A:Molecule type: DNA
A:Residues: 1-1478 <MIR>
A:Cross-references: EMBL:Z49370; MUID:91008269; PIDN:CAA89389.1; PID:91008270; MIPS:YJ
R:Cusick, M.E.
submitted to the EMBL Data Library, March 1992
A:Reference number: S27437
A:Accession: S30794
A:Molecule type: DNA
A:Residues: 602-959, 'R', 961, 'R', 963-1085, 'V', 1087, 'SLLIAHT', 1092-1094, 'RMD', 1101, 'TV'
A:Cross-references: EMBL:M88604; MUID:9172073; PIDN:AAA211179.1; PID:9172074
C:Genetics:
A:Gene: SGD:BCK1; SLK1; SSP31
A:Cross-references: SGD:S0003631; MIPS:YJL095W
A:Map position: 10L
C:Function:
A:Description: phosphotransferase; protein kinase; involved in cell proliferation
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protei
F:1173-1440/Domain: protein kinase homology <KIN>
F:1181-1189/Region: protein kinase ATP-binding motif

Query Match 4.28; Score 197; DB 2; Length 1478;
Best Local Similarity 17.78; Pred. No. 0.084;
Matches 183; Conservative 139; Mismatches 307; Indels 404; Gaps 47;
QY 34 KDSLSAERAAOAYAKEGLTTPDST-----HODSGNTEAKAGAEIYNRV-----KAA 82
Db 251 QEDISNRSSTESALSTKS-GPKSTDEKFLHSTSHQKTKSASSLYRRSFSLRGSSS 309
QY 83 KKVPLDRMPYNLOYTVYEVKNGSLI-----IPSYD-----HYHNKIFE----- 119
Db 310 SNASSAKSPSNIKLSIPARPHSIIESNLTLSKSPASPSPSPSIFRRHHKSSSESLL 369
QY 120 --WFDGLYEA-----PKGYSLDLLATVYVYEPNASHDVHKNKADQSKPDEDKEH 171
Db 370 NSLFGSGIGEAPTKPNQGHSL-----SSENTAKGK-----SKHYETNVS 410
QY 172 DEVSEPHPEDEKEN--HAGLNPSADNLKYPSTDTETEEDAEEDTTDEARIPCTPSIRQ 229
Db 411 SPLQSSILPTSDDGKLNKFKRKSQIGVSPNVAVTSQE-----TPSLKS 458
QY 230 NA-----METL-----TGLKSLLLGTKNNTNISAEVDSLLALLKESQ 267
Db 459 NSSTATLTQVADVNIPSSSSPPPIPKTANRSLEVTEDTPKIS-----STTASFKEY 514
QY 268 PAPIQCGIQGPTLPNNSLATPSPSLPNTSHEKH-----EEDGCGFDANR-- 315
Db 515 PDCINPDK-----TVP-----VPVNNOKYGVKNFLLDQKEYPLUKTKGLNDSSENYI 560
QY 316 IIAEDSGFV-----MSH-----GDSNHIF 335
Db 561 LYTNDVSYFVPLNLSKVALKSSPKESALTGLGINKNVTFTHTDQCDIGRAIPDDTLEF 620
QY 336 FKDL-----TEQIKAAQKHLEVTKSHNG--LDSLSSEHQDYPGNAKEMKOLD- 383
Db 621 LKSLFLNTSGKIYIKQDKLQKPKAPLTSENNVPLKSVKSSNRSGTSSLIASSTD 680

384 -----KIEKIAIKMOK-----GVKRESIVVNEK 410
 Db 681 VSIYSSSDITSFDEHAGSGRRYQTPSYDYDRVNTNPTEELNYWKE-----VLSHEE 737
 Qy 411 NA--IYPSGDHHDADIDEH-----PVGIGHSHSNYELFKPEGVA---KKGKNVY 459
 Db 738 NAKVFTKSPKLENLDPKSKLNPITITENESKSFVLKDKDEGTIDFNHRESY 797
 Qy 460 TGEEL-----INVNLLANSFNQNFVLANGQKRVFS-----PPPELEKL 502
 Db 798 TKPELAPKREAPKPANTSPORTLSTK-QNKPILVRASTKISKRSKPLPPOL---L 853
 Qy 503 GINMLVKLITPDGKLVKRVGKVFGEVGNIANFELDQPLPGOTFKYTIASKDYP----- 558
 Db 854 SSIPIASSSSPD-----SITSSYTPAST--HVLIPQPKYGAND 889
 Qy 559 -----EVSVDGFTVPTSLAYKM-----ASOTIYPFHAGDTYLRVNPQFAVPGK 603
 Db 890 VMRLTKDQDSTSP-SLKMQRKVRNSSTVSTNSIFYS----- 929
 Qy 604 TDALVYVDFEFGHAYLENNYKVEIKLPIPLKNGTTRTAGNKIPVTFMANAYILDNOST 663
 Db 930 -----PSLLKRGNSARV---VSSTAAADIFEENDIT 958
 Qy 664 YIVEVPILEKENQTKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSSETGNS 723
 Db 959 FADAPMPTSDSDSDSS-----SSDIIWSKKKTAPETNENKDKSDNSSTHS 1009
 Qy 724 NSTLEEVPTVDPQVKVAKFAESYGMKLENVLFNDGTYELYPGSEVKKKNADPTGEA 783
 Db 1010 DEIFYDSOTQDKMERKNT-----FRPSPEVYQNLKFP---F 1043
 Qy 784 POGNGENKPSGKSVGTGVENQTEKNKPADSI-----PEAPNEKPVKPNST 831
 Db 1044 PRAN-LDKPITTEGAS-----PTSPKSLDLSLSPKNVASSRTEPSTPS-RPVPPDSY 1094
 Qy 832 D-----NGMLNP 838
 Db 1095 EFTQDGLNGKNP 1107

RESULT 15
 T34418
 hypothetical protein F12F3.3--Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34418
 R:Fulton, B.; Wohlhmann, P.
 submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid F12F3.
 A:Reference number: Z21521
 A:Accession: T34418
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-3488 <full>
 A:Cross-references: EMBL:U80022; PIDN:AAC2585.1; GSPDB:GN00023; CESP:F12F3.3
 A:Experimental source: strain Bristol N2; clone F12F3
 C:Genetics:
 A:Gene: CESP.F12F3.3
 A:Map position: 5
 A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 4.28; Score 197; DB 2; Length 3488;
 Best Local Similarity 18.78; Pred. No. 0.29;
 Matches 108; Conservative 151; Mismatches 433; Indels 234; Gaps 37;

Qy 6 TDDEIQVAKLAGYTTEDGYFTFTSWIKKLSLSEAEARAAQAYAKEGLFPPTDHDQSG 65
 Db 714 TDSKLEAAADTKQIETETVVDKSS---KKVLKKTKESDSFISQKSETPPVPEPTKA 770
 Qy 66 NTEAKGAEATYNRV-----AAKVPIDRMPY---NLOVTVKNGSLI 106
 Db 771 ESEAQKIAEV-NKAKQKQVEDDNLKREAEVAAKKIADKLEAEANIKKTAEEVAK-- 827

Qy 107 IPSYDHYHNIKFEWFEDEGLYEAPKGYLSLELLATVKKYVVEPRNASDHVR-----NKKA 159
 Db 828 -KQEKDQKLETEFVYVSKSAEKELEK-QAQIKKAAE-----ADAVKKQKELNEKNKL 881
 Qy 160 DODSKPDDEK-----EHDEVSEPHPSDEKHNAGLNPSADNLYKSTDTDETEE 210
 Db 882 EAKKSAADKLBESAAKSKVSESVKFECKTKRAG-----BKTVOVESEP 931
 Qy 211 EAEVTTDEAIPGTPSIRQNAWETLGLKSLLLGTQDNNTIS--AEVDSLALLAKESOP 268
 Db 932 TSKTIDTKDVGATEPADETPKKIIKKKTE-----KSDSISQKSATDSSEKVKQKQD 986
 Qy 269 APIQPOIGQPTLPNNSLATPSLPINPGTSHKHEEDGYGFDANRIIAEDSEGVMSH 328
 Db 987 EPT-----KPAVSTQVTEADK-----SKQKETDEKLKDAEIAAKTKQEADEKSK 1034
 Qy 329 GDSNHYFFKCOLTEBQIIKAAOKHL-----EEVYTSINGLDSLSHSEQDYPGN----- 375
 Db 1035 LDAQEKI--KKVSEDDAARKEKELNDKLKLESEIATKASADKLKLEQQAQAKAAVEA 1092
 Qy 376 AKEMKDLKKIE-----EKIAGIMKOYGVKRESIVVNEKNAIYPSG 418
 Db 1093 AKQKQEKDEQLKIDTEAASKAAAEKLEKQAOQIKKAAGD---AVKKQKE----- 1141
 Qy 419 DHHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKVTYGEELTNVNLKKNSTFNN 478
 Db 1142 -----LDEKNKLEANKSAAGLKIEESAASK---QVVEQOAKLDAQTKATAEK 1190
 Qy 479 QNFTLANGQKRVSPFPPELEKKLINMLVLIPTDGKVLK-----VSQKV 525
 Db 1191 QT-KLEDEKSTK-----ESESKEVTDE-----KPKKKVKKKTEKSDSISQKSETSKT 1239
 Qy 526 FGEVGNIANFELDQPLPGOTFKYTIASKDYEV-----SYDGTFTVPTS 571
 Db 1240 VYESAGPSES-----ETQKVADAAARKQETDEKQKLEAEITAKSADEKSKLEAE 1289
 Qy 572 LAYKMASOTIFYPFHAG-----DTYLRVNPQFAPVPGTDALVRVDFEFGHAYLENNYK 625
 Db 1290 SKLKAAAEV-----EAAKQKQEKDEQLKIDTEAASKAAAEKL-----ELEKQSHIKKAAE 1340
 Qy 626 VGEIKLPIPLKNGTTRTAGNKIPVTFMANA---YLDNQSTYIVVEVPILEKENQTKPSI 682
 Db 1341 VDAVK-----KQKELEKQKLESEAAATKKADAELKLEOKKAAEATLIEIOKEQK--- 1393
 Qy 683 LPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSSTNSTLEEPTVDPVQEKVAK 742
 Db 1394 -----LAQEQSLRLEDEAKSAEKQKLESETKSKQTEAPKESVDEKPKKVLKKKTEK 1446
 Qy 743 FAESYGMKLENVLFNDGTYELYPGSEVKKNNM-----ADPTGEAPQNGEN 790
 Db 1447 SDSSISQKSARSATVDAAEATLESDFNLVEKKTQVKQSPDESTSATIKRDPQAQKTEBI 1506
 Qy 791 KPSENGKSVGTGVENQTEKNKPADSLPEAPNEKPVKPNSTNGMLNPENGVSDPMLDP 850
 Db 1507 SKQDDGDEKKTITDGGPP--KPEDS--EAPPKRVVKKTKQKSDSVASDASLADVSKLSD 1562
 Qy 851 ALPEAPVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEV 896
 Db 1563 DVEEKPKKVLKKKTEK-----SDSVISSETSSVDYTIKPSVEI 1600

Search completed: March 27, 2003, 11:57:55
 Job time : 63 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:12 ; Search time 27 seconds
(without alignments)
1391.761 Million cell updates/sec

Title: US-09-884-465A-332

Perfect score: 4728

Sequence: 1 MQITYTDEIQVAKLQYK.....IELRLPSGEVKKNSDFIA 906

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	204	4.3	1790	1	USOL_YEAST
2	197	4.2	1478	1	BEK1_YEAST
3	194.5	4.1	1164	1	BAG_STRAG
4	189.5	4.0	2748	1	NDM1_YEAST
5	189	4.0	1637	1	MSRP_STRAU
6	186	3.9	1658	1	YM67_YEAST
7	183	3.9	1636	1	BUD3_YEAST
8	181.5	3.8	1139	1	HMW1_MYCGE
9	180.5	3.8	1183	1	CNA_STAAU
10	180	3.8	1639	1	MSPI_PLAFW
11	176.5	3.7	2195	1	SC16_YEAST
12	176	3.7	2738	1	PCGV_RAT
13	174	3.7	3924	1	ANK2_HUMAN
14	173.5	3.7	710	1	LT78_ARATH
15	172.5	3.6	1701	1	MSPI_PLAFW
16	171.5	3.6	825	1	SW13_YEAST
17	171	3.6	1240	1	YNI1_YEAST
18	171	3.6	1630	1	MSPI_PLAFW
19	170.5	3.6	1545	1	IGA3_HAEIN
20	170.5	3.6	1701	1	MSPI_PLAFW
21	169.5	3.6	2805	1	MAPA_HUMAN
22	169	3.6	639	1	ACTA_LISMO
23	168.5	3.6	3135	1	S230_MOUSE
24	168	3.6	1612	1	FT2B_MOUSE
25	167.5	3.5	1726	1	MSPI_PLAFW
26	167.5	3.5	1726	1	MSPI_PLAFW
27	166.5	3.5	1233	1	YF16_YEAST
28	166.5	3.5	1612	1	TP2B_CRILO
29	166	3.5	1616	1	P300_MYCGE
30	165.5	3.5	1181	1	SCA2_STRPY
31	165	3.5	1849	1	IGA4_HAEIN
32	164.5	3.5	3381	1	PCGV_BOVIN
33	163.5	3.5	2869	1	RBPI_PLAVB

34	163	3.4	1466	1	SPA2_YEAST	P23201 saccharomyc
35	163	3.4	2774	1	MAPA_RAT	P34926 rattus norv
36	162.5	3.4	1766	1	SPL1_YEAST	P20134 saccharomyc
37	162.5	3.4	1781	1	AKAC_HUMAN	Q02952 homo sapien
38	161.5	3.4	1251	1	RBP2_PLAVB	Q00799 plasmodium
39	161	3.4	1167	1	SCA1_STRPY	P15926 streptococc
40	161	3.4	3122	1	POZ_MOUSE	Q61493 mus musculu
41	160.5	3.4	1095	1	NEB1_RAT	O35867 rattus norv
42	159.5	3.4	1381	1	YBE7_YEAST	P34216 saccharomyc
43	159.5	3.4	1664	1	INT1_CANAL	P33705 candida alb
44	159	3.4	1007	1	RGAL_YEAST	P39083 saccharomyc
45	159	3.4	1189	1	YJH6_YEAST	P47035 saccharomyc

ALIGNMENTS

RESULT 1
USOL_YEAST STANDARD; PRT: 1790 AA.
AC P25386; 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X2180-1A;
RX MEDLINE-91183402; PubMed-2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC '- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
COMPLEX.
CC '- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH INTRACELLULAR
MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
ER AND THE GOLGI COMPLEX.
CC '- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC '- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54378; CAA38253.1; -
DR EMBL; L03188; AB00143.1; -
DR EMBL; U53668; AB06659.1; -
DR PIR; A38455; A38455.
DR SGD; S0002216; USOL1.
DR InterPro; IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.

AIN 1 724 GLOBULAR HEAD.
725 1790 COILED COIL (POTENTIAL).
465 487 CHARGED (HYPER-HYDROPHILIC).
991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
1172 1786 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6C2B216E9FD4818 CRC64;
Query Match 4.39; Score 204; DB 1; Length 1790;
Best Local Similarity 19.79; Pred. No. 0.033;
Matches 182; Conservative 141; Mismatches 342; Indels 260; Gaps 41;
QY 67 TEAKGAEAIYNRVKAARKVPLDRPYNLYQYVVEKNGSLIIPSYD-----HYHNKFE 119
DB 650 TKTGLKDNYSRIKQFK-----DSYFSKVDNEDSILPELDEGLPKVYESTYFIQ 702
QY 120 WFDGLYKAPKYSL---EDLATYKYVYEPNADSHVKNKADODS-KPEDKEHDEVS 175
DB 703 LFNENIYRIRALSHDPEEPISKIS-PEVEKLQROCTKLKGETTSQTETSTHENLT 761
QY 176 EP-----THESDEKENHAGLSADNLYKPSDTYTE-EEADYTDAAEIPGTFSIR 228
DB 762 EKLIATWEHKEDEK---YQILNHSLSL-KENFSILETLKNVDSLSDE-----808
QY 229 QNAMETLGLASSLLGTYKNNTISAEDVDSILALLKESQAPIQGQPTLPNNSLAT 288
DB 809 -----WQLRDVLTKEKQNTALLEKYSI---HQEDSIKLEKLETI-----851
QY 289 PSPSLPMPGTSHEKEDG-----YGDANRIADESGFVMSHGD-----NHVFFK 337
DB 852 -----LSQKKAEDGINKMGKDLFALSREMQAVEENCKNLQKEKDSNVNKHET 901
QY 338 KDLTEE-----QIKAAOKHLEKVTSSNGL-----DSLSEHQDYPGNAKEMKDLKKIE 387
DB 902 KSLKEDIKAIKTEIKAINENLEEMKIQCNLSKEKEHLSKELVETKSPQSHDNVAKLT 961
QY 388 EKIAGIKQY---GVKRESIV---VKNKEI-AIIPSGDHHADPIDEK---PVGIGHSH 438
DB 962 ERLKSLANNYKDMAENSLIKAVEESNESSIQLSNLQKIDMSQENKFNQIERGSIE 1021
QY 439 SNYELFK-----PEGVAKKEGKNTYVGEELTNVNNLKN-----STFNQNFILA 484
DB 1022 KNIQKRTISDLEQKEEIIKSDSSK-----DEYESQISLLKELKLETATTANDENVNKI 1077
QY 485 NGOKRVSPFPPELE--KKGINMLVKLITPDGKVLKV-----S 522
DB 1078 SELTITRELEALAYANKLKEETKLETSE-KALKEVKEHEHLEKKEEIQLEKEAFET 1136
QY 523 GRVGEVGNITANFELDQYLPFGQTFKY--ITASQDYEVSDGTFTTPTSLAYKMASQT 580
DB 1137 KQOLNSLRANLESLEKEHEDLAQKLYEEQIANK---EROYN-----1176
QY 581 IYPPHAGDTYLRVNPQF-AVPKGTDALVRVDFPHGNAYLENNTYVGEIKLPKPLNQG 639
DB 1177 -----EISQNDLSTQENESIKKKNDLEGEV-----KAMKS 1212
QY 640 TIRTAGNIPVTMANAYLDNOSTVIVPILKKNQNDKPSILPQFK-----687
DB 1213 TSEOSN-----LKSIDALNLOIKE---LKKANETNEASLESISKVSSETVKIKEL 1263
QY 688 -----BNKAQENSKLDERVE--EPKTSEKVEKLESE-----TGNSTNSNLTVEPVTD 734
DB 1264 QDECNKEKEVSELEDKKASDKNSKYLEQKESEKIKEELDANTTELKIQLEKITNLS 1323
QY 735 PVQEK-----VAKFAESYGMKLENVLFNMDGTIELPLSGEVIKKNADFT 780

DB 1324 KAKESELSRLKKTSSERKNAEQLEKLN-----EIQKNQA---1364
QY 781 GEAPOGNGENKPSGKSVCTGVENQPTENKPADSLPEAPNEKPVPTENSTGMLNPEG 840
DB 1365 -----FEKERLLNKGSS-STTQISEKINTLEDELIRLQNELKAKEIDNTRSELEKY 1418
QY 841 NVGSDPMDLPALPEAPAVDPVQEKL 865
DB 1419 SLSNDELLE---EKQNTIKSLQDEI 1440
RESULT 2
BCK1_YEAST
ID BCK1_YEAST STANDARD; PRT; 1478 AA.
AC Q01389; P32894;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein kinase BCK1/SLK1/SSP31 (EC 2.7.-.-).
GN BCK1 OR SLK1 OR SSP31 OR LAS3 OR YJL095W OR J0906.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92186847; PubMed-1545797;
RA Costigan C., Gehring S., Snyder M.;
RT "A synthetic lethal screen identifies SLK1, a novel protein kinase
homolog implicated in yeast cell morphogenesis and cell growth.";
RL Mol. Cell. Biol. 12:1162-1178(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92104496; PubMed-1840547;
RA Irie K., Araki H., Oshima Y.;
RT "A new protein kinase, SSP31, modulating the SMP3 gene-product
involved in plasmid maintenance in Saccharomyces cerevisiae.";
RL Gene 108:139-144(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-EG123;
RA MEDLINE-92107166; PubMed-1729597;
RA Lee K.S., Levin D.E.;
RT "Dominant mutations in a gene encoding a putative protein kinase
(BCK1) bypass the requirement for a Saccharomyces cerevisiae protein
kinase C homolog.";
RL Mol. Cell. Biol. 12:172-182(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RA MEDLINE-95176706; PubMed-7871887;
RA Miosga T., Boles E., Schaaff-Gerstenschlaeger I., Schmitt S.,
Zimmermann F.K.;
RT "Sequence and function analysis of a 9.74 kb fragment of
Saccharomyces cerevisiae chromosome X including the BCK1 gene.";
RL Yeast 10:1481-1488(1994).
RN [5]
RP SEQUENCE OF 602-1104 FROM N.A.
RA Cusick M.E.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SERINE/THREONINE PROTEIN KINASE INVOLVED IN A SIGNAL
TRANSDUCTION PATHWAY THAT PLAY A ROLE IN YEAST CELL MORPHOGENESIS
AND CELL GROWTH. THIS PATHWAY SEEMS TO STARTS BY SMP3; THEN
INVOLVE THE KINASE PKC1 THAT MAY ACT ON THIS KINASE. BCK1 PROBABLY
PHOSPHORYLATES MKL1 AND MKK2 WHICH THEMSELVES PHOSPHORYLATE THE
MPK1 KINASE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE KINASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

-----NOT ANNOTATED_CDS.

DR EMBL; M4389; ; NOT ANNOTATED_CDS.
 DR EMBL; D10389; BAA01276.1; .
 DR EMBL; X60227; CA42788.1; .
 DR EMBL; X79233; CA45496.1; .
 DR EMBL; Z49370; CA489389.1; .
 DR EMBL; Z49369; CA489388.1; .
 DR EMBL; M86604; AAA21179.1; .
 DR PIR; S20117; S20117.
 DR PIR; S22285; S22285.
 DR PIR; JQ1118; JQ1118.
 DR PIR; JQ1432; JQ1432.
 DR HSRP; P24941; ICRP.
 DR SGD; S0003631; BCK1.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; P000001; Euk_pkinase; 1.
 DR SMART; SM00220; S_TK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; Phosphorylation.
 FT DOMAIN 1175..1440
 FT NP_BIND 1181..1189
 FT BINDING 1204..1204
 FT ACT_SITE 1303..1303
 FT MOD_RES 1134..1134
 FT VARIANT 1119..1119
 FT VARIANT 1120..1120
 FT VARIANT 1120..1120
 FT VARIANT 1146..1146
 FT VARIANT 1174..1174
 FT CONFLICT 59..59
 FT CONFLICT 79..79
 FT CONFLICT 264..264
 FT CONFLICT 279..279
 FT CONFLICT 703..714
 FT CONFLICT 795..795
 FT CONFLICT 802..802
 FT CONFLICT 808..808
 FT CONFLICT 903..903
 FT CONFLICT 919..919
 FT CONFLICT 960..962
 FT CONFLICT 1086..1104
 FT SEQUENCE 1478 AA; 164194 MW; D586C3A497A5BB33 CRC64;

Query Match 4.2%; Score 197; DB 1; Length 1478;
 Best Local Similarity 17.7%; Pred. No. 0.055;
 Matches 183; Conservative 139; Mismatches 307; Indels 404; Gaps 47;

QY 34 KDSLSAEARAAQAKERGLPPSTD-----HQDSNGTEAKGAELIYKRV-----KAA 82
 DB 251 QEDISNRSSTESALSPTKS-GPSKTDENFLHSTHOKTRKSASSLYRRSFISLRGSS 309
 QY 83 KVPDLDRNPNLYQYVEVANGSLI-----IPSYD-----HYHNKFE----- 119
 DB 310 SNASSAKSPNKLKISAPPHISIESNSTLTKSASPSPSPYIFRRHHKSSSSSLL 369
 QY 120 --WFDGLYE-----PKGISLEDLLATVYVYPRNASDVRKKNKADQDQSPDEK 171
 DB 370 NSLFGSGIGEEAPTKPNPGHSL-----SSENLAKG-----SKHYTNWS 410
 QY 172 DEVSEPTHPESDEKEN--HAGLNPSADNLYKPSDTEETEEAEADTDEAIPGTPSI 229
 DB 411 SPLKQSLPTSDDKGLWKNFKRKSQIGVPSPTNVAIVTSQE-----TPSLKS 458

QY 230 NA-----METL-----TGLKSSLLGTCKDNTTISAEYDLSLALKESQ 267
 DB 459 NSSTATLVQADVNIPSPSSPPPIPTANRSLVSTEDTPKIS-----STTASFKEYY 514
 QY 268 PAPIOGPOIGQPTLPNNSLATPSLPINPGTSHKHK-----EEDGYGFDAHR-- 315
 DB 515 PDCINPDK-----TVP-----VPVNNQKYSVKNFLDQKFKPLKTLGLNDSENKYI 560
 QY 316 IIADESSEFV-----NSH-----GDSNHYF 335
 DB 561 LVTKDNVSFVPLNLKSVAKLSFKESALTGLGINKNVTFTMTDFCDIGAAIPDITL 620
 QY 336 FKKDL-----TEEQIKAAQKHLKEVYKTHNG--LDSLSHEDQYDPGNKAKKOLD- 383
 DB 621 LAKSLFNTSGKIYIKDQMKLQKPKAPLTSENNVLKSVKSKSMRSGTSSLASTDD 680
 QY 384 -----KKIEKIAGIMQY-----GVKRSIIVNKEK 410
 DB 681 VSIVTSSSDITSFDEHAGSGRRYPQTPSYVYDRVSNTPTEELNWNKIE---VLSHEE 737
 QY 411 NA--IYPSGDHHDADPIDEHK-----PVGIGHSHSNYELFKPEGVA---KKEGK 459
 DB 738 NAPKRVFKTSPKLELNLPDKSKLINIPTITENESKSSQVLRKDEGTIDFNHRES 797
 QY 460 TGEEL-----TNVNLKKNSTFNQNFNLANGOKRVSF-----FPPELEK 502
 DB 798 TPELAPKREAPKPPANTSPQRTLSK--ONAPIRLVRASTKISRSKRPPLPQL---L 853
 QY 503 GINMLVKLITPDGKLVKRVGKVFGEVGNIANFELDQPYLPQOTFKYTIASKDYP-- 558
 DB 854 SSPIEASSSPD-----SLTSSYTPAST--HVLIPQPYKGAND 889
 QY 559 -----EVSDGTFTVPTSLAYKM-----ASQTIYFPHAGDTYLVRNPQFV 603
 DB 890 VMRLKLTQDQSTSTSP-SLKMOKRVNRSNSTVSTNSIFYS----- 929
 QY 604 TDALVRVDFEFGHAYLENNYKVGKIKPLKNOQTTRTAGNKIPVTFMANAYLDNQST 663
 DB 930 -----PSPLLKRGNSRV-----VSSTSAADIFENDIT 958
 QY 664 YIVEVPILEKQTKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKISGTGNS 723
 DB 959 FADAPMFSDSDSDSS-----SSDDIWSKKKTAPNTNENKDKSDNSSTHS 1009
 QY 724 NSTLEVPVTPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVKKKADFTGEA 783
 DB 1010 DEIFYDSQTDKNERKMT-----FRPSPVYVYQNLKFP---F 1043
 QY 784 POGNGENKPSKNGKYSTGTVENQNTENKPADSL-----PEAPNEKPVKPNENST 831
 DB 1044 PRAN-LDKPITTEGIAS-----PTSPKSLDLSLSPKNVASSRTEPSTPS-RPV 1094
 QY 832 D-----NGMLNP 838
 DB 1095 EFTQDGLNGKNKP 1107
 RESULT 3
 ID BAG_STRAG STANDARD; PRT; 1164 AA.
 AC P27951;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Iga FC receptor precursor (Beta antigen) (B antigen).
 GN BAG.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1311;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.

STRAIN-LA239;
MEDLINE-91312121; PubMed-1857207;
Jarlstrom P.G., Chatawal G.S., Timmis K.N.;
"The Iga-binding chain antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions.";
Mol. Microbiol. 5:843-849(1991).
[2]
IDENTIFICATION OF IG-LIKE DOMAIN.
MEDLINE-37035265; PubMed-8880921;
Bateman A., Eddy S.R., Chothia C.;
"Members of the immunoglobulin superfamily in bacteria.";
Protein Sci. 5:1939-1942(1996).
-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).
-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/> or send an email to [license@sib-sib.ch](mailto:license@sib-sb.ch)).
EMBL; X59771; CAA42442.1; --
PIR; S15330; FCSOAG.
InterPro; IPR004829; Surface_antigen.
InterPro; IPR001899; Gram_pos_anchor.
InterPro; IPR003599; Ig.
Pfam; PF00746; Gram_pos_anchor; 1.
ProDom; PD153432; Surface_antigen; 2.
SMART; SM00409; Ig; 1.
TIGRfams; TIGR01167; LPXTG_anchor; 1.
TIGRfams; TIGR01168; YSRK_signal; 1.
PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
Immunoglobulin domain.
SIGNAL 1 37
CHAIN 38 1135 IGA FC RECEPTOR.
PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).
DOMAIN 434 534 IG-LIKE DOMAIN.
DOMAIN 199 438 IGA-BINDING (POTENTIAL).
DOMAIN 439 826 IGA-BINDING (POTENTIAL).
DOMAIN 827 945 PRO-RICH REPEATS.
SITE 1132 1136 LPXTG SORTING SIGNAL (POTENTIAL).
MOD_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SEQUENCE 1164 AA; 65DE94AF720A5474 CRC64;
Query Match 4.1%; Score 194.5; DB 1; Length 1164;
Best Local Similarity 19.4%; Pred. No. 0.052;
Matches 166; Conservative 130; Mismatches 316; Indels 245; Gaps 40;
QY 166 DDKEHDEVESETHP---ESDKEHAGLNPSADNLYKPSDTETEEAEATTTDEAIP 222
DB 43 DSVKTEVAAPKSLTGLKSL---LGTKDNNTISAEV-----DSLLALKESQAPAIQG 273
QY 223 GTPSTRONAKMETLGLKSL---LGTKDNNTISAEV-----DSLLALKESQAPAIQG 273
DB 103 DPGKREKQLQWKNLKNVDVNTILSHQKNEFKYKIDETNDSALLEL----- 151
QY 274 POIGOPTLPNNSLATPSPSLPINSPTSEKHEEDGYGFDPANRIIAEDSGFVNSHGDG--- 331
DB 152 -----ENQFNETNRLHLK---QHEVEKDKKA-KQKTLKQSDTKVDSIDKREL 198
QY 332 NH-----YFFKDLTEQIQAQKHLLEVTSNGLDLSLSSHEQDYPGNKEMKDLK 384
DB 199 NHOKSOVKASQKGTNEDKDSMLKIEDIRKQAQADKEDAE-----VKVREELGK 252
QY 385 KTEEKIAGIMK---QYGVKRETVVAKKNAIYPSGDHHAADI----- 426
DB 253 LPSSTAGLDQEQEHVKET---SSEENT---QKVDHEYANSLQNAKSLLELDKAT 306

QY 427 DEHKPVGIGHSHSNYELFKPEGVAKKEGKNVY--TGEELTNVNVLLKSTFNQNFLLA 484
DB 307 NEQATQVKQKLENAQKLEIKIPLIKETNVKLYKAMSESLQVEKELKHSEANLEDLVA 366
QY 485 NGOKRVSFPFPPLEKLGINMLVKLITPDGKLVKSGVKGVGNI-ANFELDQPYL 543
DB 367 KSEIYR-----EYEGKL--NQSGL--PELQLEEEAHSKLQVVEDFRKKEFTSEQVT 417
QY 544 PGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYKMASQTIFFPFHAG 588
DB 418 PKKVRDLAANNQKIELTVSPENITVEGEDVKFTVTA---KSDSKT---TLDIFS 470
QY 589 DTYLRVNPQPAVPGKTDALVRVDFEFGHAYLENNYKVGSEIKLPIPKLNOGTRT----- 643
DB 471 DLLTKYNPSVS-----DRISTNYKTNT---DNHKAIEITIKNLKLSNDSOTVTLKAKD 519
QY 644 -AGNKIPVTMANAYLDNQSYIVEVPILEKENOTDKPSILPOFKRKAQENSKLDEKV- 701
DB 520 DSGNVVEKTF-----TITVQKKEK-----QVPTPEOKDSKTEKVP 557
QY 702 EEPRTSEKVEKELSETGN-----STNSNTELEVPTV-----DPVOE 738
DB 558 QEPKSNQKQLQELIKSAQOQLEKLEKAIKELMEQPEIPSNPEYGIQKSIWESQKEPIQE 617
QY 739 KVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGVKKNMADTGEAPQG 786
DB 618 AITSFKKIIGDSSSKYTYEHFNKYSDFMNYQLHAQM-----EMLTRKYVQVNNKYPDN 672
QY 787 NGENKPSENGKVS-----GTVENOPT-----NK-----PADSLPEAPNEKP 824
DB 673 AEIKKIFESDMKRTKEDNYGSLNDALKGFPEKFLYFPFNKIKQIVDDDKKVEQDOPAP 732
QY 825 VKPENSTDNGLNPEGNVGDPMLEDALEA-----PAVDPVOEKELEKFTASYGLG 875
DB 733 I-PENSE-----MDQAKERAKIAVSKYMSKVLGDVGHQKKNNSKIVD 775
QY 876 L-----DSVIFNMD 884
DB 776 LFELEAIKQOTIFDID 792
RESULT 4
NUMI_YEAST STANDARD; PRT; 2748 AA.
AC Q00402; (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nuclear migration protein NUM1.
GN NUM1 OR YDR150W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28383 / FL100;
RX MEDLINE-92079907; PubMed-1745235;
RA Kormanec J., Schaaff-Gerstensschlaeger I., Zimmermann F.K.,
RA Percec D., Kuentzel H.;
RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the highly repetitive 313 kDa NUM1 protein.";
RL Mol. Gen. Genet. 230:277-287(1991).
CC -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR ENVELOPE.
CC -1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH FLANKING DOMAINS OF THE TANDDEM REPEATS.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF115379; AAD09131.1; -
 CC InterPro; IPR001899; Gram_pos_anchor.
 CC Pfam; PF00746; Gram_pos_anchor; 1.
 DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
 DR TIGRfam; TIGR01168; YSRK_signal; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 KW Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor;
 KW Repeat; Signal.
 FT SIGNAL 1 48 POTENTIAL.
 FT CHAIN 49 1601 METHICILLIN-RESISTANT SURFACE PROTEIN.
 FT PROPEP 1602 1637 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 1301 1584 141 x 2 AA TANDEN REPEATS OF D-[SAG].
 FT SITE 1598 1602 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1601 1601 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 4.0%; Score 189; DB 1; Length 1637;
 Best Local Similarity 19.8%; Pred. No. 0.15;
 Matches 208; Conservative 118; Mismatches 415; Indels 308; Gaps 47;

QY 6 TDEIQVAKLAGYTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPSTDQDQSG 65
 DB 124 TTEAPKAECTDRVTEAPKAEET--DRATTEAPKAEETDKATEA---PKTEETDKA 178
 QY 66 NTE-----AKGAAIYRVKAAKVPDLDRPYNQYTVVEKNGSLIIPSYDHYHNK 117
 DB 179 TTEAPAAEETSKAATEAPKAEETSKAATEAP---KAEETKATERAPKTEETDRYE 235
 QY 118 FEWFDEGLYAPAGYSLEDLLATVYVYPRNADVRKADQDSKPEDKEHDEVSPP 177
 DB 236 TE-----EAPKAE-----TSKATEKAPKAEET--NKVETEAPAEETNKAATEE 280
 QY 178 THP-ESDEKENHAGLNPS-----ADNLYKPSDTETEETEAED----- 214
 DB 281 TPAVEDTNKASNSNAQPSERTQVVDVTAQDLYKKSEVTEAEKAEIEKVLPRDISNLSN 340
 QY 215 -----TTDAEIPGV--PSTRONAMETLGLKSL--L-----LGTKDNN 250
 DB 341 EEIKKIALSEVLKETANKENQAPRATRSVSNARTNINYSATLRAAQAQDTVYKGTG 400
 QY 251 TISAEVDSLALAKESQAPIQGQPTLPNNSLATPSP-----SLPI 295
 DB 401 NPTAHGDIHKTVEE--PPNEGTLAFATNPNPTGTGALRYNDKIDFNKDTITVPV 458
 QY 296 NPGTSHEKHEEDGYFDANRIIAD---ESGFVMSHGDNSHHFFPKDLTEEQIKAAQKHL 352
 DB 459 ANNNQGNNTGADGNGFMFTQGGQDFLNOGGILDRKGMANASGFKIDTAYNNVNGKVDKL 518
 QY 353 BEVTSNGLDLSLSSHQDYPGNAKEMKLDKIEEKIAGIKQYGVKRESIVVNRKE-- 410
 DB 519 DADT--NNLSQIAAKVGVGTFKNGAD-----GVTHQVG--QNALTKDKPV 563
 QY 411 NAIYPSGDHHAADPIDEHKVPVIGSHS-----NYELPPEGVAKKGNKYVTE 462
 DB 564 NKIIYADNTNHLND-----GPHGQRLNDVVLNYD-----AATSTTATYAGK 606
 QY 463 ELTVNVLNLSNFNNQFTLANGQKRVST--FPPELEKLGINMLVKLITPDGKVL 519
 DB 607 TWKATDGLDGLDKSQKYNFLITSSHQNRYSNGIMRTNLEG-----VVTITQADLID 659
 QY 520 KVSQKVEGEGVGNIANFELQDPLPGQTFYTIASKDYPEV-----SYDG--TFTVPTSLA 573
 DB 660 DV--EVTQKPIPHKTIREFDTLEPGS-----PDVIVKQGEDGKTTTPT--- 703
 QY 574 YKMASOTIFPFHAGDLYLRVNFQFAPVKG--TDALVRVF-----DEFHGNAYLENN 623
 DB 704 -KVDPDT-----GDVVERGEPTTEVTKNPVDEIVHFTPEEVPQGHKDEFPN----- 749
 QY 624 YKVGKILPPLKNGQTRTAGNKIPVTFMANAYLDNQSYIIVVEPILK-----E 674

DB 750 -----LPI-----DGTEEVPG-----KPGIKNPETGEVVTTPVDDVTKHGPAGE 789
 QY 675 NQTDKPSILPOFRANKAQENSKLDKVEEPTKSEKKEKLSGTGNSNSTLEEV--- 730
 DB 790 PEVTKKEIPPEKKREFNPDLKPGEEKV---TOGGTGEKTTTTPTTINPLTGERKGE 845
 QY 731 PTVDVPOEKVAKFAESYGMKLE-----NVLFNMGGTIELYLP-----SGEVIK 774
 DB 846 PTVETVEPDEITQFGEEVPOGHKDEFPNLPIDGTEE--VPGKPGIKNPETGEVVT 903
 QY 775 NMAFTGEAPGN-----GENKPSNGKVSCTGV----- 803
 DB 904 PVDDVTKHGPAGEPEVTKKEIPPEKKREFNPDLKPGEEKVTOGGTGEKTTTTPTTINP 963
 QY 804 -----ENQPT---ENKPADSLPEAPNEK-----PVKPNSTDN-----GMLNP 838
 DB 964 LTGEKVEGEPTTEVTKPEVDEITQFGEEVPOGHKDEFPNLPIDGTEEVPKPGIKNP 1023
 QY 839 E-GNVGSDPMLDPALEAPAVDPVQEKLE 866
 DB 1024 ETGEVVTTPVDDVTKHGPAGEPEVTKKEE 1052

RESULT 6

YMG7_YEAST STANDARD; PRT; 1658 AA.
 AC Q03661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 187.1 kDa protein in GUAL-ERG8 intergenic region.
 GN YMR219W OR YMR261.13 OR YMR959.01.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OK NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; 249809; CAA89934.1; -
 DR EMBL; 249939; CAA90190.1; -
 DR SGD; S0004832; YMR219W.
 KW Hypothetical protein.
 SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.9%; Score 186; DB 1; Length 1658;
 Best Local Similarity 19.1%; Pred. No. 0.21;
 Matches 217; Conservative 164; Mismatches 389; Indels 366; Gaps 53;

QY 6 TDEIQVAKLACK---YTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPSTD- 60
 DB 583 SEEQIPKDTGENTNNKTDGRDLSSSVTEIEVKVSEKKLDG-----STEELVPLSDT 638
 QY 61 -----HQDSGNTGAEAGAEIYNRVKAARKVPLDRPYNQYTVVEKNGSLIIPSYDH 112

Db 639 TINSSGLNEDSIYSLDADAISENL---TDVPLMEIKTKPKYEVVISESVYSSTSYED 695
QY 113 -----YHNKFEWDEGLYAPKGYSLDLELATVKKYVEB----- 147
Db 696 NTVAMPQVETSPFNDPNSLND--YB-----KKHDLKSLALAPAFYKKADEFVE 749
QY 148 -----RNASHRVKKNKADQSK--PDEKHEHVESEPTHPSEDEKENHAGLNPASDN 197
Db 750 AGVTKSCLSTSGHTNIPHTSKETQVSLDSESTENVTEENGTGDKNKNQKNPFGVAN 809
QY 198 LYPKSTDTETEBAEDTDE-----AETPGPSIRQNAEMETLGLKSSLLGKTGN 249
Db 810 -----STDKSTEDNTDEYFSAINTNTVGTSSC-EDIETASNEEENLRICEKM 859
QY 250 NTISAEVDSLLALLKESQAPIQGPGIOTPLPNSLAPSPSLPNCPTSHEKHEEDY 309
Db 860 N--EAMSSGDECVQND-----GSKT--QTSFSTDSFD---NFQESNDTEFST 904
QY 310 GFDA-NRIIADESGFVMSHGDSNHYFFKKDLTEQIKAAKHEEVKTSNGLDLSLSS 368
Db 905 KYKVRNSDLEDDDES-----LKLTKAEB--VDKLDEB-----ESEDY 941
QY 369 EODY-----PCNAK-----EMKDLKTEBK-----AGIMKQYGVK 401
Db 942 EQDYADPEFGNDEGSNENYKTKKDTLGIPEPEKVKVHEEETLFEANVSSSYNVQN 1001
QY 402 ESI---VNVKEKNALITPSGDHSH-----ADPIDEKPVGIGHSHSNYELF 444
Db 1002 KMHDTVINGEQAQ-NYBAGERKYIYQNTDTEAHISIIERIDEN---AIG---NNKEI- 1053
QY 445 KPEGVAKKEGKRYVTGEBLTNVN--LKNSTFNQNFLE-----NGKRVSF 492
Db 1054 -PERSCEKTHNEVLFERRATTENTKALENTNMHDQVQSCSDSDRQDQSTAEKNVGS 1112
QY 493 SPFPELEKLGIML--VKLTPDGVLKVSQ--KVFG-----EGVGNIANFELQDPLPG 545
Db 1113 SAKHNDIRVSSEIESVPLAPESDRSIFSPIRVIGAVKGVGVYDV-----A 1164
QY 546 QPKFYIASKDPEVBYDGTFTVPTSLAYKASQTFYFPFHAGDTVLRYNPQPAVPGTD 605
Db 1165 ESF---VKIDVMSDDNDVIGD-----YNQDIFNSNTDASVNMKSYSKERSD 1215
QY 606 -----ALVRVDFEFGHAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLN 660
Db 1216 EDEAVILGGVTAEAHND--GNNSRVINI-----DPTNGAYEEDSEVFQV-KDK 1264
QY 661 QSTYIVEVPIL-----KENQDKPSI----- 682
Db 1265 ENLHSEPLVGELOEQHFKEKDHSENSEEEFTIYGDITSANIHSNAPDDIKRQOLLN 1324
QY 683 -----LPQFRKKAQENS-KLDEKVEPKTSEK-----VEKEKLSGTNS 721
Db 1325 LSDLENYSQLIEDSRGNQDESVNTSREDLTFEASVNEKYAGAEEDTFSELDIS 1384
QY 722 TSNSTLEVPVDPVQEK-----VAKFAESVGMKLE-----NVLNMDGTIELYIP 767
Db 1385 IQHPEHEEDLNNQERSTEELNSEPEAELEYEIEGTETATAASSKKNDDRQRGNIP 1444
QY 768 SCGVIKKNAD---FTGEAPOGNGENKPNKGVSTGTVENOPTENKPADSLPEAPNEK 824
Db 1445 STDLSPDPSDEEVTDSYPYSENITAE-----KSAPTSPEVETFTDTPNEVP 1495
QY 825 VK-----PENSTDNGLMPEGVNGVSDPMLDPA 851
Db 1496 MEINDEIPATLLEKHDKNTVSVLDRSEHLSHVDNPHONSINKVNEGEPP----- 1550
QY 852 LEBAVD-----PVQEKLEKTASTGLDLSVIFNMGTIELRLPSGEVKKNLS 902
Db 1551 --EHOAVDIPKVKVEEQEEMPSK-----SVLEQKPSMELINDKSS 1591
RESULT 7
BUD3_YEAST

ID BUD3_YEAST STANDARD; PRT: 1636 AA.
AC P25558; P25557; P87007;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bud site selection protein BUD3.
GN BUD3 OR YCL014W OR YCL14W/YCL13W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBL_taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95247824; PubMed=7730410;
RA Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT "Role of Bud3p in producing the axial budding pattern of yeast.";
RL J. Cell Biol. 129:767-778(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Staveva L.I.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVISIONS.
RA Newes H.-W.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U17580; AA86315.1; -
DR EMBL; X59720; CAA42346.2; -
DR PIR; S19341; S19341.
DR PIR; S19340; S19340.
DR SGD; S0000520; BUD3.
DR InterPro; IPR000219; RHOGEF.
DR SMART; SM00325; RHOGEF; 1.
KW Cell cycle.
SQ SEQUENCE 1636 AA; 184717 MW; 9E4E46BA5C3A3F69 CRC64;
Query Match 3.9%; Score 183; DB 1; Length 1636;
Best Local Similarity 19.0%; Pred. No. 0.29;
Matches 186; Conservative 146; Mismatches 316; Indels 332; Gaps 50;
QY 5 YTDDETOVAKLAGKYTDEGYIFD-----TSWIKKDSLSEAEAAAAQAY 48
Db 690 HDDKHIEV-----TSNIVFTIINQIAIEIPICFSSLNSSMAKDLVCYENLNLE 741
QY 49 AKEKGLTPSTDHQDSGNTAKGA---EAIYNRVKAARKVPLDRMPTNYQYTYVEVNGSL 105
Db 742 HQLEEVKHPSTDEHRAVNSKLSGASDFDATHKKRSGYGTITFR-----SYTSLDKDS-- 794
QY 106 IPSYDHYHNKFEWDEGLYEAPKGYSLDLE---ATKYVVEPR-----NAS--D 152
Db 795 --PSGDNSNVTK-----ETKEILPVKPTKSSKKKPREIOKTKTNASKAE 837
QY 153 HVRKNKADO-----DSKPDDEKHEHVESEPTHPSEDEKENHAGLNPASD 196
Db 838 HIEKKAPNKGKGFVGLKNVFGSKSKSPVQVRPKISQRHPKSPVK----- 886

Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.
Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.:
The minimal gene complement of Mycoplasma genitalium.*;
Science 270:397-403(1995).
[2]
SEQUENCE OF 721-847 FROM N.A.
RP STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RT Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III,
RT *A survey of the Mycoplasma genitalium genome by using random
RT sequencing.*;
RL J. Bacteriol. 175:7918-7930(1993).
CC -I- STABILIZES THE SHAPE OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC FUNCTIONALIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -I- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC MEMBRANE (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See http://www.isb-sib.ch/annou-
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39712; AAC71534.1; --
DR EMBL; U02261; AAD12527.1; --
DR TIGR; MG312; --
KW Cytoadherence; Structural protein; Complete proteome.
DR
KW SEQUENCE 1139 AA; 130531 MW; 0011D3288C3DD856 CRC64;
SQ

Query Match 3.8%; Score 181.5; DB 1; Length 1139;
Best Local Similarity 19.7%; Pred.No.0.21;
Matches 161; Conservative 118; Mismatches 342; Indels 197; Gaps

QY 146 EPRNASHVRKKNADQSDPKDEKHDEVSEPTHPESDEKENHAGLNPSADNLKPSTDT 205
DB ||| | :
DB 18 EPENIFGLDTGKSTVEEDPNIKVAYD-----ADNGYYIAFNKETGYIDPYDGT 68
QY 206 EETEEAEADTTDAEIPGPSIRQNAMETLTGLKSSLLGTDKNNNTISA-----E 255
DB ||| | :
DB 69 EYDISQLFDEN-----GNPFVDFQEENDYLK---YVGNDPDSYDENGEMWNSGYFE 119
QY 256 VDSLALLAKESQPAPIQ-----GPQICQPIPLNNSLATPPSPPLPNCISHERHEEDG 308
DB ||| | :
DB 120 NDQWIS--TKESQPTDENYGFSDDLPEVPKQESVEDNYGFDN-DLP--PEVKQPSVEDN 175
QY 309 YGFDANRI-----IAEDSGFVMSHGDNSHHFFFKDLTTEQIIKAQAQHLEVKTSHLGDSL 365
DB ||| | :
DB 176 YGFD-NDLPEVPKQES--VDQPSDDDIFAKQP-TDENYGFDNDLPEVPKQESVVDP 231
QY 366 SS---HEODYPGNAREKMDLKKIEEKIAGIMKQYGVKRESIVVNNEKKNAIITPSPGDHHA 423
DB ||| | :
DB 232 SSDDHFAKQPESTITDSYSFSDLQPQ-----TLDPQLSDHRVQ 269
QY 424 DPIDEH---KPVGIGHSHSNVELFKPEEGVAKEGKNVTGEELNVNLLKNSTFNQN 480
DB ||| | :
DB 270 YNFDBHEELXPV-AEQNNYQV-----GFDQVQAN-LDNNEFIOTPAEKRYTTFDESQ 321
QY 481 FTLANGQRVRSFPFPLEKLGINMLVKLIT-PDGKVLEKVSCKVFGSGVGNIANFELD 539
DB :
DB 322 AQVVD-----SYQLPIDTGQQDTFTSSSFETQTVQEQDVQNSEV----- 362
QY 540 QPYLPGGTQKYTTASKDYPEVSYDGTFTVPTSLEYKMASQTIFYPPHAGDTYLRVNPQFA 599
DB ||| | :
DB 363 -----NQOFFPETKEPVLESFNKDQVETS---DLNESNLYSENKDA----- 405
QY 600 VPKGTDALRVF-----DEPHGNAYLE--NNYKVGEEKLPKPLNGOTTTRTAGN 646


```

Db 406 --TNDLSNSETASDSDVHYSEKSEPHDKFGS-----DLQSNNSNLE 457
Qy 647 KIPVTFMA-----NAVLDNQSYI--VEVPILLEKEN---QTKPSILPQPKRKAQENSKL 697
Db 458 SEPVKFNSETAPDAHFSEQVQVQYDIYQNEELKPTLDQPSDDYFAKQPTDITG 517
Qy 698 DEK-----VEEPTSEKVEKELSTGNTSGNSTLEEVPTDP-----VQ--- 737
Db 518 DNDLPEVKQESVVDQSSDDHFAKPESTDSFSDLPQ--PTLQPSLDDHVQNF 576
Qy 738 -----ERKVAFAESYGMKLENVLMGDTIELYLPSEGVIRKKNMADFTGEAPQNGENK 791
Db 577 DHHEELKPAVEENQYQVGFQVQANLNEEIQ--PTAE--KEVTTDFESKQAQ----- 627
Qy 792 PSNGKVTGTVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPGNVGSDPMDLPA 851
Db 628 VYDSQLPIDT--DOODQTTSSSFETQPTVEQFQVNVSEVNDQKPE--ITKEPVLESS 683
Qy 852 LEEAPAVDQVEKLEKFTASVGLGLDSVFNMDGTIEL 889
Db 684 FNKQDVVE-----TGNNTNKLQKFDIQSDNKITI 712

RESULT 9
CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FDA 574;
RX MEDLINE-92165839; PubMed-1311320;
RA Patti J.M., Jonsson H., Guss B., Switalaki L.M., Wiberg K.,
RA Lindberg M., Hoecek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalaki L.M., Wiberg K.,
RA Lindberg M., Hoecek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN-FDA 574;
RX MEDLINE-94032261; PubMed-8218209;
RA Patti J.M., Boles J.O., Hoecek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE-97475225; PubMed-9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoecek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M81736; AAA20874.1; --
DR PDB; 1AMX; 24-JUN-98.
DR InterPro; IPR001899; Gram_pos_anchor.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1154
FT PROPEP 1155 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT SITE 1151 1155
FT MOD_RES 1154 1154
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E57D76 CRC64;

Query Match 3.8%; Score 180.5; DB 1; Length 1183;
Best Local Similarity 21.2%; Pred. No. 0.24; Indels 261; Gaps 46;
Matches 198; Conservative 122; Mismatches 351;

Qy 46 QAYAKEK--GLTPSTDHQDQNTAKGAEIYVNRKAAKVPKLDMPY--NIQYTVVEKN 102
Db 300 QEHGKEVNGKSPNHTVHNANAGIEG--TVKGLKVLKQDKQKAPIANVKFSLSKD 357

Qy 103 GSL-----IIPSYDHYHNKFEWDEGLY-----EAPGYSLEDLLATVKTVEPR 148
Db 358 GSVKDNQKEIEITDANGIANIKA--LPSGDYILKEIEAPRYTFD-----402

Qy 149 NASDHVRNKADQDSKPDDEK-----EHDEVSPHTPESDEK--ENHAGLNPSA-DNL 198
Db 403 -----KDKEYPFTMKDQDNOGYTTIENAKAIEKTDVSAQKRWEGTKVKPYIFKL 455

Qy 199 YKPSDTTEEEAEEDTT--DEAEI-----PGTPSIQONAMETLTGLKSLSLGTLKDNNTI 252
Db 456 YK-----QDDNQNTTPVDKAEIKLEDGTTKVTW-----SNLPENDKNGRAI 497

Qy 253 SAEVDSLLALKESOPAIQGIQOPIPLNNSLATPSPSLPI--NPGTSHHEHEDEGTF 311
Db 498 KYLVKEVNAQGEDTTPGY-----TKENGLVVTNTEKPIETTSISGEKRVWDKQ 549

Qy 312 DANRTIADESGFVMSHSDSNHYFPFKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQ 371
Db 550 DGR--PERKSVNLLANG-----EKVT-----LDTSETNWK 580

Qy 372 YPGNAKEMKDLDDKIE--EKIAGIMKQYGVK-----RESIVYVK--EKNAIIPSGDH 420
Db 581 Y-----EFDLPKYDEGKIEVTVEDHVKYDVTDDINGTTITNKYTPGTSATVTKNWD 635

Qy 421 HHADPIDEHKPVGIGHSHSNYELFPBEGVA-----KKEGKV 458
Db 636 NNQ--DGRKPTTEI-----KVELY--QDGKATGKTALNESNNWTHWTGLDEKAGQOV 686

Qy 459 -YTGELTNVNLKSTFNNOFTLANQKRVSFPPPELEKLGINMLVKLITPDGKV 517
Db 687 KTVSELTKV-----KGYTHVDNDM--GNLIVTNKPTTETTSISGEKRVWDKQDGR 740

Qy 518 LEKVSQKVFEG-----VGNIAFELDQPYLP-----QTEFYTIAS---KDYPEVSD 563
Db 741 PEKSVNLLADEKVKTLDTSETNWKYEFKDLPKYDEGKIEVTVEDHVADY--TTDIN 799

Qy 564 GTFTVPTSLAYKMAQSOTIFYPPHAGDTYLRV-----NPQFAVPKGTDLAVRVDE-- 613
Db 800 GT-----TINKYTPGTSATVTKNWDNNQDGRKPTTEIKVELYQDGRAT 845

```

-----FHGNAYLENNYKVEIKLPIPKLNO-----GTTTAGNKIPVTFMANAYL 658
 846 KXTALLNESNWTHTWGLDEKAGQVKTYVELTKVGTTHVDNDGNLIVTKYT 905
 659 DNQSTYIIVEPILKQNTOK-----PSILPOPKRKA-----QENSKLDE 699
 906 PETTSISGEKVDKNDQGRPEKVSNNLLANGKVKTLDTSETNKVEFKDLPKYDE 965
 700 --KVEEPTSEKVEKELSETGNSSTSL--EEVPTVD-----PVQEKVAK 742
 966 GKKEIYTVTEHDVYDITTINGTITTKYTPGETSATVTKNWDNNQDKRPTTEIKVEL 1025
 743 FAESYGMLENLVNMDQTEIYLPSPGVKKNMADEF-GEAPQGNENKPSNGKYSTG 801
 1026 YQDKATCTAILNESNNWHTWGLDEKAGQVKTYVDELTKVNGTYTHVDNDGNL 1085
 802 TVENOPTENKPADSL-PEAPNEK--PVKPENS 830
 1086 IVTKYTPKPKNKPIYPEKPKDKTPTKPDHS 1117

RESULT 10
 MSPL_PLAFW STANDARD; PRT; 1639 AA.
 AC P04933;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (P195).
 DE (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Wellcome).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID:5848;
 RN [1]
 RP MEDLINE-86014355; PubMed-2995820;
 RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
 RA Nicholas S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
 RA Freeman R.R.;
 RT Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites.*;
 RL Nature 317:270-273(1985).
 RN [2]
 RP REVISIONS.
 RA Holder A.A.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X02919; CAA26676.1; .
 DR PIR; A24594; A24594.
 DR InterPro; IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1639 MERZOITE SURFACE PROTEIN 1..
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 768 768 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 3.8%; Score 180; DB 1; Length 1639;
 Best Local Similarity 19.1%; Pred. No. 0.4;
 Matches 150; Conservative 125; Mismatches 282; Indels 228; Gaps 37;

QY 149 NASDHYRKNKADQSDPDEKEDH-----EVSEPHPESEKENHA----- 189
 DB 112 NPSD-----NSSDSDAKSYADLKHVRNYLLTILKELYPQLFDLTNMLTCDNIHFKYL 167
 QY 190 --GLNPSADNLKYPSTDETEEEAE--TTDEAEIPGTPSIRQNAETITGLKSSLLG 245
 DB 168 IDGYEETINELLYKLNFFDLRAKLVNDVCANDYQIPFNKIRANELDVL---KKLVFG 223
 QY 246 -TKDNTISAEVDSLLALKESQAPIQGPOIGOPTLPNNSLATPSPSLDINPGTSHEKH 304
 DB 224 YRKPLDNKDNVGMEDYIKKN-----KKTEN-----IN-----ELI 256
 QY 305 EEDGYGDANRIATAEDSGFVMSHSDSNHYFFKKDLTEEQ--TKAQKHLEEVKTSN-- 360
 DB 257 EESKKTIDKNKATKEEEKKKLYQAQYDLSIYNKQLEAHNLISVLEKRIIDTLAKKENIK 316
 QY 361 -GLDLSLSHEDQYPGNA---KEMKOLDKRIE---KIAGIMK-----Q 396
 DB 317 ELLDKINEINPPANSNGTPTLLDKNKKEIEKEIEKIATKFNIDSLFTDPLELE 376
 QY 397 YGVR-----RESIVVKEKNALIIIPSGDHHADPIDEKHPVIGHSNHYE 442
 DB 377 YLREKKNIDISAKVETKSTEPNEYPNGVTPSYNDINNALNELNSFG-----D 428
 QY 443 LFKPEEGVAKKGVVTGEELTNVNLKNSTPNQNFNLIANGQKRVSVSFPELEKKL 502
 DB 429 LINPFD-YTEPSKNITDNERKAFINEIEK-----IKIEKKIESDKKSYEDRSK 479
 QY 503 GINMLVKLITPDGKVLKGVGEGVGNIANFELDQPYLPQGTFFYTTASKDYPEVSY 562
 DB 480 SLNDITKEYE---KLLNEIYDSKENNNI-DLTNPE---KMGGRYSYKV-----EKLTH 526
 QY 563 DGTFTVPTSLAYKMASQTIFFPHAGDTYLVRVNPQFAVPKGTDLVRVDFEHGNAYLEN 622
 DB 527 HNTFA-----SYE-----NSKHNLKLTALKYKNEYSLRNIVVEK 562
 QY 623 NYKVGELKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKQNTDKP-- 680
 DB 563 ELKY--YKLLSKIE-----NEIB-TLVENIKKDEQLF--EKKITDENKPDKIL 609
 QY 681 --SILPOPKRKAQENSKLDKVEEPTSEKVEKELSETGNSSTSLTELEVPTVDVQVE 738
 DB 610 EVSDIVKQVQVKVLLMKNKIDELKKTOLILKNVLEKHNHVPNSYKQENKQEPYLLVKK 669
 QY 739 KVAKFAESYGMLENLVNMDQTEIYLPSPGEVI---KKNMADFTGEAPQGNEN-KPS 793
 DB 670 EIDK-----LKVFMPKVESLINEEKNI-----KTEGQSDNSEPS 704
 QY 794 ENGVSTG-----TVENOPTENKPAD-----SLPEAPNEKPKVP---E 828
 DB 705 TEGEI-IGQATKPGQAGSALGDSVQAQAEQQAQPPVPVPEAKAQVPTPPAPVN 763
 QY 829 NSTDN 833
 DB 764 NKTN 768

STANDARD; PRT; 2738 AA.
084; 008592; 088564; 09R1K4;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Versican core protein precursor (Large fibroblast proteoglycan)
(Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
hyaluronate-binding protein) (GHA) (Fragments).
CPSG2.
Rattus norvegicus (Rat).
Mus musculus; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_Taxid-10116;
[1]
SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
(ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
STRAIN-Wistar Kyoto;
STRAIN-Wistar Kyoto;
Medline-99327053; PubMed-10397680;
Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
Wright T.N.;
"Versican/PG-M isoforms in vascular smooth muscle cells.";
Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
[2]
SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
STRAIN-Wistar Kyoto;
Medline-98308094; PubMed-9642104;
Millev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
Margolis R.K., Margolis R.U.;
"Differential regulation of expression of hyaluronan-binding
proteoglycans in developing brain: aggrecan, versican, neurocan, and
brevican.";
Biochem. Biophys. Res. Commun. 247:207-212(1998).
[3]
SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
TISSUE-Kidney;
Medline-98094159; PubMed-9434070;
Pytko C., Kristensen P., Ostergaard P.B., Otural P.S., Romer J.;
"Proteoglycan expression in the normal rat kidney.";
Nephron 77:461-470(1997).
[4]
SEQUENCE OF 2535-2738 FROM N.A.
STRAIN-Sprague-Dawley; TISSUE-Lung;
Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
"Molecular cloning and characterization of two developmentally
regulated genes in rat lung.";
Submitted (SEP-2000) to the EMBL/GenBank/DBSJ databases.
FUNCTION: May play a role in intercellular signaling and in
connecting cells with the extracellular matrix. May take part in
the regulation of cell motility, growth and differentiation. Binds
hyaluronic acid.
SUBCELLULAR LOCATION: Secreted; extracellular matrix.
ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and
Vint; are produced by alternative splicing.
TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
but not in glomeruli.
DEVELOPMENTAL STAGE: Disappears after the cartilage development
(by similarity).
SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
SIMILARITY: CONTAINS 2 LINK DOMAINS.
SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AF062402; AAC40166.1;

1423 TVD--LTENRDIENVNSTMVSDLPTWEPKLSKPE-----VNEPKQKSGS 1467

RESULT 13

ANK2_HUMAN	STANDARD;	PRT; 3924 AA.
ID	ANK2_HUMAN	
AC	Q01484; Q01485;	
DT	01-APR-1993 (Rel. 25, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DT	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).	
GN	ANK2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
GN	[1]	
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).	
RC	TISSUE=Brain stem;	
RC	SEQUENCE=Brain stem;	
RC	MEDLINE=91302466; PubMed=1830053;	
RA	Otto E., Kunimoto M., McLaughlin T., Bennett V.;	
RA	"Isolation and characterization of cDNAs encoding human brain	
RT	ankyrins reveal a family of alternatively spliced genes.";	
RL	J. Cell Biol. 114:241-253(1991).	
RL	[2]	
RN	SEQUENCE OF 463-495 FROM N.A.	
RP	Carpenter S.;	
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	
RL	[3]	
RP	SEQUENCE FROM N.A. (ISOFORM 1).	
RC	TISSUE=Brain stem;	
RC	MEDLINE=94075409; PubMed=8253844;	
RA	Chan W., Kordeli E., Bennett V.;	
RA	"440-kD ankyrinB: structure of the major developmentally regulated	
RT	domain and selective localization in unmyelinated axons.";	
RL	J. Cell Biol. 123:1463-1473(1993).	
RL	[4]	
RN	SEQUENCE OF 463-495 FROM N.A.	
RP	MEDLINE=92009921; PubMed=1833308;	
RA	Tse W.T., Menninger J.C., Yang-Peng T.L., Francke U., Sahr K.E.,	
RA	Lux S.E., Ward D.C., Forget B.G.;	
RT	"Isolation and chromosomal localization of a novel nonerythroid	
RL	ankyrin gene.";	
RL	Genomics 10:858-866(1991).	
CC	-!- FUNCTION: Attach integral membrane proteins to cytoskeletal	
CC	elements. Also bind to cytoskeletal proteins.	
CC	-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE	
CC	PRODUCED BY ALTERNATIVE SPLICING.	
CC	-!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL	
CC	CELLS THROUGHOUT THE BRAIN.	
CC	-!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES	
CC	AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE	
CC	AND FUNCTION (POTENTIAL).	
CC	-!- SIMILARITY: CONTAINS 23 ANK REPEATS.	
CC	-!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; X56957; CAA0278.1; -	
DR	EMBL; X56958; CAA0279.2; -	
DR	EMBL; Z26634; CAB42644.1; -	
DR	EMBL; M37123; AAA62828.1; -	
DR	PIR; S14533; A15333.	
DR	PIR; A39643; A39643.	
DR	PIR; B39643; B39643.	
DR	PIR; S14569; S14569.	
DR	HSP; P42771; IDC2.	

89 RMPYNQYTVFV-----KNGSLIIPSYDHYHNIKFEW-----FDEGLYEAP-----129
1521 RVEDLEKNEIILRSCTCTDESSVQSSRGLVEEWIVDSIEEARQKAPLEITE 1580
130 -----KGYSDLELLATVYKYPERN-----ASHVRK-----156
1581 YPCVEVRIDKEIKGKEDKSTGLVNTLTDLCVLPKPEQLQTVQDKAGKCEALAVGR 1640
157 --NKADDSKPDZ-----DREHDE---VSEPTHESEDE---KENHAGLNPSADNLYKPESTD 205
1641 SSEKEGKDIPDETQSTQKQKSLGKIKKPVRRKLEKQKKEGQASAKAEALAKGSS 1700
206 EETEEEAEDTDEAEIPGTPSIRQNAEMETLGLK-SLLGLTKDNNTTISAEVDSLLALLK 264
1701 EES-----LGEDPGL-----APEPLTVKATSPLEETPIGSIKDKVKALAKRVE 1745
265 EQ-----PAPIOGPO-----IGQPLPNNSLATPSPS-----292
1746 DEQGRSKLPVVKGEDVPKKTTHRPHPAASPFLKSERHAPGSPSPKTERHSTLSSAK 1805
293 ---LPINPGTSHKHEEDGVGFANRIIAEDSESGFVMSHG-----DSNHYFFKDLT 341
1806 TERPPVSPSPSKTEKHPSPVSPSATERHSPASSSSKTEKHPSPSPSKTERHSPVSTKT 1865
342 EQ--IKAAQKHLKEEVKTSNGLDLSLSSHODYPGNKAKEMKDLKKEEIKIAGIMKOYGV 399
1866 ERHPPVSPSGTKDRPPVSPSG--RTEKHPPVSPSGT-----EKRLPVSPSGRTDKH-- 1915
400 KRESIVYKNEKNAIYPGDDHHDAPID-EHKPGVIGHSNHYLEKFP-----446
1916 QPVSTAGTEKHLVPSPSGTEKQPPVSPSPSKTERHSPVSPSKTERHSPVSTKT 1975
447 ---EGVAK-----KEGNKVTYGEELTVNVLNKNSTFNQNFNLANGOKRVSFSPPELE 499
1976 GLFEHKSAAQKQPOEKGKRVVEKEKGP--LTQREAOKTENQTIKRGOR---LPVGTAE 2030
500 KKLGINMLVLIITPDGKVLKVGKVGEGVGNANTANFELDQPLPGQTFKTYIASKOYPE 559
2031 SKRGVR-----VSSIGVKKEDAAG-----GKEKVLSHKIPVP-----QSVPE 2068
560 VSYDGTFTVPTSLAYKMASQITFYPPFHAGDTYLRVNPQFVAPKGTDLVLRVDFEFGNAY 619
2069 BESHRESEVPKE---KMADE-----QGDMDLQISDPDKTSTDFSEVIK-----QE 2110
620 LENNYKVGEEKLIP-----KLNGTTRTAGNKIPVTFMANAY-----LDNOST 663
2111 LEDNOKYQFRLSEETERAKQLHLDQVLTSPFNTTFLDYMKDEFPLALSLQSGALDSS- 2169
664 YIVEVPILKENQTDKP--SILPQKRNKAQENSKLDEKVEEPT-----SEKVEKEK 714
2170 -----ESLKNEGVAGSPGSLMEGTPOISSEESYKHEGLAETPETSLSFSPPKSEQ 2224
715 LSETGNSTSNSTLEEV-----PTVDPV---QEKVAKFAESYGMKLENVLNFMNDGTIEL 764
2225 TGETESTKTTETTEIRSEKEHPTTKDITGSGSEERGATVTEDETSETSES--FQKATLIG- 2281
765 YLPSGEVTKNNMADTG-----EAPGNGENKTPSNGKRVSTG--TVENOPTENKPADS 815
2282 --SPKDTSPKRDCTGSCSVALAKETPTGLTEEAACDEGORTFGSSAHKTQTDSEAQES 2339
816 -----LPEAPNEKVPKPNSTONGMLNPNNGVSDPMLDPA 852
2340 TATSETKALPLPEA-----SVKTDGTES---KPOGVINSPOGLEAL 2380

RESULT 14
L778_ARATH
ID LT78_ARATH STANDARD; PRT: 710 AA.
AC Q06738; Q04981;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

Query Match
Best Local Similarity 19.5%; Pred. No. 2.4;
Matches 197; Conservative 125; Mismatches 400; Indels 286; Gaps 47;
35 DLSLEAERAAQAYAKEGLTPSTHQDSNGTEAGAEIYNRVRAAKKVP-----LD 88
1469 DLLSEVSEMKQDLIKWTAILT-----TVSDKA-----GSIKVKELVRAAEEPEPEIVE 1520

Low-temperature-induced 78 kDa protein (Desiccation-responsive protein 29A).
 LTI78 OR LTI140 OR RD29A OR COR78 OR AT5G52310 OR K24W7.4.

Arabidopsis thaliana (Mouse-ear cress).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 eurosids II; Brassicales; Brassicaceae; Arabidopsi.

NCBI_TaxID=3702;

[1]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia; TISSUE=Leaf;

MEDLINE=93192524; PubMed=8448363;

Nordlin K., Vahala T., Palva E.T.;

"Differential expression of two related, low-temperature-induced

genes in Arabidopsis thaliana (L.) Heynh.;"

Plant Mol. Biol. 21:641-653(1993).

[2]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia;

MEDLINE=94143472; PubMed=8310052;

Yamaguchi-Shinozaki K., Shinozaki K.;

"Arabidopsis DNA encoding two desiccation-responsive rd29 genes.;"

Plant Physiol. 101:1119-1120(1993).

[3]

SEQUENCE FROM N.A.

McLarnay B.K., Gilmour S.J., Thomashow M.F.;

Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.

[4]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia;

MEDLINE=30181125; PubMed=10718197;

Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,

Tabata S.;

"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence

features of the regions of 3,076,755 bp covered by sixty p1 and TAC

clones.;"

DNA Res. 7:31-63(2000).

[5]

SEQUENCE OF 357-710 FROM N.A.

STRAIN=cv. Columbia; TISSUE=Leaf;

MEDLINE=91322497; PubMed=1830821;

Nordin K., Heino P., Palva E.T.;

"Separate signal pathways regulate the expression of a low-

temperature-induced gene in Arabidopsis thaliana (L.) Heynh.;"

Plant Mol. Biol. 16:1061-1071(1991).

CC -1- TISSUE SPECIFICITY: ACCUMULATES RAPIDLY IN LEAVES AND ROOTS.

CC -1- INDUCTION: BY WATER STRESS OR ABSICISIC ACID (ABA), AND MOSTLY BY

CC LOW TEMPERATURE.

CC -1- SIMILARITY: BELONGS TO THE LTI78/LTI65 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.ch).

CC EMBL; X67671; CAA47903.1; .

CC EMBL; D13044; BAA02376.1; .

CC EMBL; L22567; AAA32775.1; .

CC EMBL; AB019226; BAB10528.1; .

CC EMBL; X57600; CAA40826.1; .

CC PIR; S16273; S16273.

KW Multigene family; Repeat.

FT DOMAIN 63 67

FT POLY-ASP. REPEATS OF P-[MV]-G-F-G-[DS]-

FT 2 X 14 AA REPEATS OF P-[MV]-G-F-G-[DS]-

FT E-S-G-A-E-L-E-K.

FT 1-1.

FT REPEAT 303 316

FT 1-2.

FT REPEAT 357 370

FT 3 X 15 AA REPEATS OF [DN]-[FS]-P-[STV]-R-

FT S-H-[DE]-[FL]-D-[LM]-K-[NT]-E-[ST].

FT REPEAT 317 331

FT REPEAT 336 350 2-2.
 FT REPEAT 398 412 2-3.
 FT DOMAIN 510 600 5 X 5 AA REPEATS OF [FV]-[ADT]-[EST]-
 [KW]-L.
 FT REPEAT 510 514 3-1.
 FT REPEAT 532 536 3-2.
 FT REPEAT 550 554 3-3.
 FT REPEAT 579 583 3-4.
 FT REPEAT 596 600 3-5.
 FT DOMAIN 638 641 POLY-GLY.
 FT DOMAIN 648 696 2 X 23 AA REPEATS.
 FT REPEAT 648 670 4-1.
 FT REPEAT 674 696 4-2.
 FT CONFLICT 216 216 S -> P (IN REF. 2).
 FT CONFLICT 491 491 E -> V (IN REF. 2).
 FT CONFLICT 514 514 L -> H (IN REF. 5).
 SQ SEQUENCE 710 AA; 77856 MW; 9C6C8ACAE6BDF334 CRC64;

Query Match 3.78; Score 173.5; DB 1; Length 710;

Best Local Similarity 19.9%; Pred. No. 0.26;

Matches 153; Conservative 82; Mismatches 271; Indels 263; Gaps 32;

QY 284 PINPQTSHEKHEEDCYGPDANRIIAEDSGFVMSHGDSNHFFKKDLTEOIKAAOKHLE 353
 DB 8 PLN---THQHPPE-----VEHENGATKMKFR-----VKRAKFK 41
 QY 354 EVKTSNGLDLSHSHEQDYPGNNAKEMKDKKIBKIKAGIMKQYGVKRESIVNVNKK--- 410
 DB 42 NSLTRHG---QSNHEQDH--DLVEEDDDDDLEPEVIDAPGVTKPRETNVPAEEIIP 96
 QY 411 -NAIYPSGDHHADPIDEHKPV-----GIGH---SHS-----NYLFKPEGV 450
 DB 97 PCKTVPVPSYDTPK-TESPVQEAHYGHADPAHSVTTFTSDKEEKRDYPIHHPLSEL 155
 QY 451 AKKENKVTYTGELNVNLLKNSFFNQNT-----LANGKRVSPSPF----- 495
 DB 156 SDRESETHESLNTVPVLLSGTSDTSTFAPSGDDEYLDQKRVNVTPTILEESAV 215
 QY 496 -----PELEKLGINKLVKITPDGK----- 516
 DB 216 SDYLSGVSNYSQKVTDPKTEETGGVPEIAESFG-NNEVTDESPDQKPGQFERDLSTRKE 274
 QY 517 -----VLEK-----VSGKVGEGVGNIANFANLPGLPGLPGTKFYTIASKDYP 558
 DB 275 FKEFOODFDSVLGKDSPAKFPGESGVVFPVGGDSGSAELE-----KDFP 319
 QY 559 EWSYDCTFTVPTSLAYKMASTIFYPFHAGDTYLRVNPQ-FAPVPGTDALVNVDFEFGN 617
 DB 320 TRSHDFDMKTEGMDTNSPSRSHPEPLKTESGNDKNSPMGFGSESGAE-LEKEFDQKND 378
 QY 618 AYLENNYKVEIKLPIKLNQQTTRTAGNKIPVTFMANAYLDNQSTIYVEVPI----- 670
 DB 379 G--RNEYS-----PESDGGIGAPLGGNFPVR-SHELDLKNESDIDKDVPTGDFGEPD 427
 QY 671 -----LEKENOTD-----KPSILPQFKRK 690
 DB 428 FLAKRPGYGEASEEDKFPARSDDVEVEFELGDDPTEYLDQFSPSLSPKDERDEKESR 487
 QY 691 AQENSKLDERVEEPKTSYKVEK-EKLSETGNTSTNLTLEEVTVTD-----PVOEKVAKFA 744
 DB 488 DDFEETRDENTEPEKQSTYTEKFPASMLGYSGEIPVGDQTVAGTVDEKLTVPNEKDOETE 547
 QY 745 ESTGMKL-----ENVFNMDGTIELYLPSEGIKKKMAADFTGAPGNGE 789
 DB 548 SAVTKLPISGGGSGVEEQGEDKSVSGRDYVAEKLTTTEE--DKAFSDNVAEKLOIGE 605
 QY 790 NKPSKNGKVTGTVENOPTENKPADSLPEAPNEK-----PVKPE 828
 DB 606 EEKKE---TTTKEVEKISTE-KAASEGEAVEEVEKGGGKMGVGRKGFGGGATDEVKPE 661
 QY 829 N-----STDNGMLNPGNVGSDPMLDPALEAP-----AVDPVQEK 865
 DB 662 SPHSVEAPKSSGFWGGGATEEVKPKSPHSVEESPSQSLGSTVVPVQOKEL 710

Best Local Similarity 18.6%; Pred. No. 0.94;
Matches 193; Conservative 157; Mismatches 334; Indels 351; Gaps 51;

Y	15	LAKYTTDEG-----YIFDTSWIKKSLSEAEAA-----AAQAVAKKGLT	55
b	701	VAASETTDEGSHSTHLSQSGETEVEETEVEVTHGTTTITLPPKESAPKEVAV	760
Y	56	PPSTDRDQSGTEA-----KGAEIYVNRVKA	83
b	761	ENSIEHKNSNDNSQALTKVYLKDLDEFTKSYICHYILVNSNSMDQKLLLEV	820
Y	84	KV-----PLDRMPYNLQYTVKNGSLIPSYDYHVNKFKFEWDEGLYEA	138
b	821	KELKSCDPLDLL-FNIQ-----NNIPAMTSLYDSMNNDQLHFFPE-LYQ	866
Y	139	ATVKYVYEPNADSVRKNKADQSKPDDEKHEDEVSEPTHPESDEKENHAGLNP	198
b	867	-----YYLHKLEENHIKLLLEQ-----KQITGTSSTSPGNTTVN-TA	908
Y	199	YKPTDTEETEERADTDEAIEPTGPTSRQAMETLTLGLKSSLLGLTKDNNIT	258
b	909	THSNSQOQSNASSTNTQNGAVSSGPAVVEESHDLPTVL-----SISNDL	956
Y	259	LLALKESQAPIQGOIGOPTLPNNSLATPSPPLNPCTSHKHEEDGYGFDAN	318
b	957	IVSLL-----NLG-----NKTVPNP-LTIST-TEMEKPYEN-----	989
Y	319	EDESGFVMSHGDNSHYFFKRD-----LTEQIKAAQKHLKEEVKTS-----	361
b	990	NNDYF-----NDDIKQPVKSNKVTGLTGTQKNALNDEIKKLDLQI	1045
Y	362	LDSLSSHEQYDGNKEMKO-----LDKKIEKIAIMKQGVKRE-SIVVKN	414
b	1046	LDRLFNKKELGODKMOIKKLLKLEKSLNSLNPHNVQZNVFFNKKKEAET	1105
Y	415	-----YPSGDHHDADIDEKHPVGI--GHSNHYELFKPEGVAKK	453
b	1106	ENTLENKILLKHYKGLVKNYGE---SSPLKTLSEVSIOTEDNYANLEK	1162
Y	454	EGNKVYTGEE-----LTVNVLKNSTFNQNFILANGQKRVSFPPELEK	505
b	1163	LNDNLHLGKKLSFLSSGLHLITEL- EVIKNNY- ---GNSPSENKK--	1210
Y	506	MLVKL---ITPDGKVLKSVKGVGNGVNIANFELQPYLPQGTFKTYI- ---	557
b	1211	EALKSYENFFPEAKVTVTVTP-----PODVTPLSVRVSGSGSGSTKEE	1255
Y	558	PEVSYDGTFTVPTSLAYKASQTIFFPHAGDTYLRVNPQFAPVKGTD- ---	607
b	1256	TQIPTSG-----SLLTELQVQVQNTYDEEDSLWLPFGESEDNDEYLD	1309
Y	608	VRYDFEFGNAYLENNYKVGKIPKIPKLNQGTTRTAGNKIP---VTFMNA	656
b	1310	SVTMDNILSG--FENEYDVIYK-PL---AGVYSLKKQIEKNIITFNLND	1362
Y	657	-----YLD-----NQSTYVEVPLEKENQTDKPSILPQFRNK	690
b	1363	KKRYFDVLESLDMQFKHISSEYIIE-DSFKLLNSEQKNTLLKSYIKES	1421
Y	691	AGNSKLDEK-----EETPKSEKVEKELSTGNSNSTLEEYVTPVQEKV	742
b	1422	AQEGISYIEKVLAKYKDDLESIKKVEKEKFPSPPTTTPS-----PA	1476
Y	743	FA-----ESYGMKLENVLFNN-----DGTIELY	765
b	1477	FLPFLNIETLYNNLYNKIDYLLINKAKINDCNVEKDEARVKITKLSDL	1536
Y	766	LPSGEVKKMADFTGEAPQNGENKPSNGK-VSGTYVENQPT-----	817
b	1537	-----KNTNFEAKKLLINDTKDKMLGKLLSTGLVGNFPNTIISK	1588
Y	818	EAPNE--KPVKPPENS	830

MSPI1_PLAFM STANDARD; PRT; 1701 AA.

AC P08569; (Rel. 08, Created)

DT 01-AUG-1998 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Merzoite surface protein 1 precursor (Merzoite surface antigens) (P190).

GN MSP-1.

OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_Taxid-70153;

RN SEQUENCE FROM N.A.

RP MEDLINE-88011243; PubMed-3079521;

RA Tanabe K., Mackay M., Goman M., Scaife J.G.;

RT "Allicic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum."

RL J. Mol. Biol. 195:273-287(1987).

RN REVISIONS TO 1403; 1569 AND 1629.

RA Tanabe K.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-115 FROM N.A.

RX MEDLINE-86136024; PubMed-3004972;

RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U., Stunnenberg H., Bujard H.;

RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merzoites: studies at the genetic level."

RL EMBO J. 4:3823-3829(1985).

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (POTENTIAL).

CC -1- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X05624; CAA29112.1; -

DR PIR; A26868; A26868.

DR PIR; B25120; B25120.

DR InterPro; IPR000561; EGF-like.

DR Pfam; PF00008; EGF; 1.

DR Malaria; Merzoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.

KW SIGNAL

KW CHAIN

FT SIGNAL 1 19

FT CHAIN 20 1701

FT CARBOHYD 110 110

FT CARBOHYD 239 239

FT CARBOHYD 470 470

FT CARBOHYD 536 536

FT CARBOHYD 607 607

FT CARBOHYD 802 802

FT CARBOHYD 899 899

FT CARBOHYD 919 919

FT CARBOHYD 965 965

FT CARBOHYD 991 991

FT CARBOHYD 1089 1089

FT CARBOHYD 1196 1196

FT CARBOHYD 1588 1588

FT SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 3.6%; Score 172.5; DB 1; Length 1701;

us-09-884-465a-332.rsp

Fri Mar 28 09:21:11 2003

Db 1589 ISQHOCVKQCPENS 1603

Search completed: March 27, 2003, 11:55:11
Job time : 44 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:13 ; Search time 103 Seconds
(without alignments)
1812.414 Million cell updates/sec

Title: US-09-884-465A-332
Perfect score: 4728
Sequence: 1 MQITYTDEIQVAKLQYV.....IELRPSGEVKKNLSDFA 906

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3357	71.0	1039	16 Q9ANY1	Q9any1 streptococc
2	1307.5	27.7	839	16 Q9ANY2	Q9any2 streptococc
3	946.5	20.0	819	16 Q97QW9	Q97qm9 streptococc
4	946	20.0	844	2 Q9AG74	Q9ag74 streptococc
5	935.5	19.8	819	2 Q9ANY3	Q9any3 streptococc
6	636	13.5	802	16 Q97QW8	Q97qm8 streptococc
7	636	13.5	816	2 Q9AHY9	Q9ahy9 streptococc
8	233	4.9	825	2 Q9JG75	Q9jgt5 streptococc
9	230	4.9	825	16 Q99XV4	Q99xv4 streptococc
10	223.5	4.7	822	2 Q9ZHG7	Q9zhg7 streptococc
11	215.5	4.6	1524	10 Q8RYN2	Q8ryn2 oryza sativ
12	213.5	4.5	5458	5 Q9U459	Q9u459 plasmodium
13	208	4.4	2647	5 Q9U4X0	Q9u4x0 plasmodium
14	202.5	4.3	2269	5 Q96223	Q96223 plasmodium
15	201.5	4.3	1078	5 Q963T1	Q963t1 plasmodium
16	201.5	4.3	1873	2 Q9Z4N7	Q9z4n7 enterococcu

17	201	4.3	1790	3 Q07380	Q07380 saccharomyc
18	201	4.3	2004	16 Q97QP7	Q97qp7 streptococc
19	198	4.2	1420	5 Q9Y0V8	Q9y0v8 drosophila
20	197	4.2	3484	5 P91257	P91257 caenorhabdi
21	195.5	4.1	1139	5 Q97237	Q97237 plasmodium
22	195	4.1	1698	2 Q9LC00	Q9lc00 staphylococ
23	195	4.1	2910	10 Q9FND5	Q9fnd5 arabidopsia
24	194.5	4.1	1134	2 Q99051	Q99051 streptococc
25	193.5	4.1	1125	16 Q98PM9	Q98pm9 mycoplasma
26	193	4.1	1245	2 Q9L470	Q9l470 staphylococ
27	192	4.1	1871	10 Q9SRD5	Q9srd5 arabidopsia
28	191	4.0	3111	5 Q9VH10	Q9vh10 drosophila
29	190.5	4.0	2747	5 Q9BJX9	Q9bjx9 plasmodium
30	190	4.0	1596	11 Q62766	Q62766 rattus norv
31	190	4.0	4688	16 Q9PQ08	Q9pq08 ureaplasma
32	189.5	4.0	2748	3 Q9VC00	Q9vc00 drosophila
33	189.5	4.0	2768	5 Q9PPE5	Q9ppe5 ureaplasma
34	189.5	4.0	5005	16 Q9PPE5	Q9ppe5 ureaplasma
35	189.5	4.0	5327	5 Q76891	Q76891 drosophila
36	189	4.0	1596	11 Q921F7	Q921f7 rattus norv
37	188.5	4.0	1793	5 Q9W596	Q9w596 drosophila
38	187	4.0	2151	5 Q9NG79	Q9ng79 trichomonas
39	185.5	3.9	1983	16 Q9CJ55	Q9cj55 lactococcus
40	185.5	3.9	2735	5 Q76416	Q76416 caenorhabdi
41	185	3.9	1733	2 Q9KI14	Q9ki14 staphylococ
42	185	3.9	2485	5 Q96134	Q96134 plasmodium
43	184.5	3.9	1964	2 Q59947	Q59947 streptococc
44	183.5	3.9	1795	2 Q9LC39	Q9lc39 staphylococ
45	183	3.9	865	2 Q9FDJ1	Q9fdj1 streptococc

ALIGNMENTS

RESULT 1

ID	Q9ANY1	PRELIMINARY;	PRT; 1039 AA.
AC	Q9ANY1;		
DT	01-JUN-2001 (TREMblrel. 17, Created)		
DT	01-JUN-2001 (TREMblrel. 17, Last sequence update)		
DT	01-OCT-2001 (TREMblrel. 18, Last annotation update)		
DE	Pneumococcal histidine triad protein E precursor (Hypothetical protein SPI004).		
GN	SP1004.		
OS	Streptococcus pneumoniae.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Streptococcaceae; Streptococcus.		
OX	NCBI_TaxID=1313;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=21101045; PubMed=11159990;		
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,		
RA	Dormitzer M., Degan R., Brehm Y.A., Warren P., Lathigra R.,		
RA	Langermann S., Koenig S., Johnson S.;		
RT	"Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis."		
RL	Infect. Immun. 69:949-958(2001).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=TIGR4;		
RC	MEDLINE=21357209; PubMed=11463916;		
RA	Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,		
RA	Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,		
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,		
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,		
RA	Holtzapple E., Khouri H., Wolf A.M., Otterbeck T.R., Hansen C.L.,		
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,		
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,		
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,		
RT	"Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."		
RL	Science 293:498-506(2001).		
DR	EMBL; AF318956; AAK06761.1; -		


```
AAK75121.1;
Query Match 71.08; Score 3357; DB 16; Length 1039;
Best Local Similarity 72.36; Pred. No. 1.1e-153;
Matches 687; Conservative 40; Mismatches 95; Indels 128; Gaps 12;

QY 7 DEIOVAKLAGKTTEDGYIF-----DTS-----WIKDLSSEARAAQA 47
DB 168 NENAVARSGRYTTNDGYNPADIEDTGNAYIVPHGGHYHYPKSDLSASELAAKA 227
QY 48 YAKEGLTP-----PSTDHQDSGNTKGA-----EALYNRVKAARKVPL 87
DB 228 HLAGNMQPSQLSYSTASDNTQSVAGSTSKPANKSENQSLKELYDPSAQRVSES 287
QY 88 DRMPYNLQYTVVKNGLIIPSYDHYHNKIFWFDGLYEAAPKGYSLDLATYKYVYP 147
DB 288 DGLVDPKAKIISRTNGVAIPHGDHYHFPYS-----KLSALEEKIARM---VPI 334
QY 148 RNASHVKNKADQDSKDEKDEHDEVSEPHPSDEKENHAGLNPSADNLYKPTDTEE 207
DB 335 SGTGSTVSTN-----AKPNVY-----SSLGSSLN---PSSLTTS 367
QY 208 TEEAEDTTDEAIEPGTSPSIRQNAETLTGLKSSLLLTGKDNNTISAEDVSLALLKESQ 267
DB 368 KE-----LSSASDGYIFPKD-----IVEETA 389
QY 268 PAPI-----QGPOIGOPTLPNNSLATPGSPSLPINFPTGTSHEKHEEDGYGFDANRI 316
DB 390 TAYIVRHGDHPIYIPKSNQIGOPTLPNNSLATPGSPSLPINFPTGTSHEKHEEDGYGFDANRI 449
QY 317 IADSSGFMVNSHGDHNYFFKDKLTDEEIQKAAKHLEVKTSNGLDLSLSHEODYPCNA 376
DB 450 IADSSGFMVNSHGDHNYFFKDKLTDEEIQKAAKHLEVKTSNGLDLSLSHEODYPCNA 509
QY 377 KEMKLDKRIEKIAGIKQVGRKRESIVVANKENAIYPSGDHHDADPIDEHKPVGIGH 436
DB 510 KEMKLDKRIEKIAGIKQVGRKRESIVVANKENAIYPSGDHHDADPIDEHKPVGIGH 569
QY 437 SHSNTELPKPEGVAKGKNKYTGEEELTNVYNLKNSTFNQNTLANGOKRVSFSPFP 496
DB 570 SHSNTELPKPEGVAKGKNKYTGEEELTNVYNLKNSTFNQNTLANGOKRVSFSPFP 629
QY 497 ELEKLGIMLVKLTTPDGKLVKSGVFGEGVGNIAFELDQPLPGQTFKYTIASKD 556
DB 630 ELEKLGIMLVKLTTPDGKLVKSGVFGEGVGNIAFELDQPLPGQTFKYTIASKD 689
QY 557 YPEVSYDGTFTVPTSLAYKMASQTIFYPFHAGDTLYRNPQFAVPKGTDALVRVDFEFG 616
DB 690 YPEVSYDGTFTVPTSLAYKMASQTIFYPFHAGDTLYRNPQFAVPKGTDALVRVDFEFG 749
QY 617 NAYLENNYKVGKILPIPKLNGTTRTAGNKIPVTFMANAYLDNOSTIYVEVPILEKENQ 676
DB 750 NAYLENNYKVGKILPIPKLNGTTRTAGNKIPVTFMANAYLDNOSTIYVEVPILEKENQ 809
QY 677 TDKPSLTPOFKKKAQENSKLDKEVEEPTSEKVEKELSETGNTSNTLEEVPVDPV 736
DB 810 TDKPSLTPOFKKKAQENSKLDKEVEEPTSEKVEKELSETGNTSNTLEEVPVDPV 869
QY 737 QEKVAKFAESYGNKLENVFNMDGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSNG 796
DB 870 QEKVAKFAESYGNKLENVFNMDGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSNG 929
QY 797 KYSTGTVENQPTENKPADSLPAPNEKPVKPNSTNDGMLNPEGVNGSDPMLDPALEAP 856
DB 930 KYSTGTVENQPTENKPADSLPAPNEKPVKPNSTNDGMLNPEGVNGSDPMLDPALEAP 989
QY 857 AVDPVOEKLEKFTASVGLGDSVIFNMDDTIELRLPSGEVYKKNLSDFTA 906
DB 990 AVDPVOEKLEKFTASVGLGDSVIFNMDDTIELRLPSGEVYKKNLSDFTA 1039

RESULT 2
Q9ANY2 PRELIMINARY; PRT; 839 AA.
ID Q9ANY2
AC Q9ANY2;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TEMBLrel. 18, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHTD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacilliales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21101045; Pubmed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barran P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.,
RT Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.
RL Infect. Immun. 69:949-958(2001).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; Pubmed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., C.M.;
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.
RL Science 293:498-506(2001).
DR EMBL; AF318955; AAK06760.1;
DR EMBL; AE007403; AAK75120.1;
DR TIGR; SP1003;
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 27.78; Score 1307.5; DB 16; Length 839;
Best Local Similarity 81.78; Pred. No. 4.1e-55;
Matches 263; Conservative 2; Mismatches 6; Indels 51; Gaps 4;

QY 2 QITVTDDEIOVAKLAGKTTEDGYIFD-----TSWIKDLSSEAR 42
DB 518 QITVTDDEIOVAKLAGKTTEDGYIFDPRDITSDGDAYTPHMTSHWIKDLSSEAR 577
QY 43 AAAQAYAKEGLTTPSTDHQDSGNTKGAERAIYNRVKAARKVPLDRMPYNLQYTVYKN 102
DB 578 AAAQAYAKEGLTTPSTDHQDSGNTKGAERAIYNRVKAARKVPLDRMPYNLQYTVYKN 637
QY 103 GSLIIPSYDHYHNKIFWFDGLYEAAPKGYSLDLATYKYVVE-----PR-----NAS 151
DB 638 GSLIIPSYDHYHNKIFWFDGLYEAAPKGYTLEDLATYKYVVEHPNRPNSHDNGFNAGS 697
QY 152 DHRKNKADQDSKDEKDEHDEVSEPHPSDEKENHAGLNPSADNLYKPTDTEETEE 211
DB 698 DHRKNKADQDSKDEKDEHDEVSEPHPSDEKENHAGLNPSADNLYKPTDTEETEE 757
QY 212 AEDTDEAIEPGT-----PSIRQNAETLTGLKSSLLLTGKDN 250
DB 758 AEDTDEAIEPGT-----PSIRQNAETLTGLKSSLLLTGKDN 817
QY 251 TISAEDVSLALLKESQAPAIQ 272
```

```

Db 818 TISAEVDSLALLKESQAPAIQ 839
|||||
RESULT 3
Q97QM9 PRELIMINARY; PRT; 819 AA.
AC Q97QM9
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Conserved domain protein.
GN Sp1174.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE-21357209; PubMed-11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwion M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., E.K.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL; AF007418; AAK7283.1; -
DR TIGR; SP1174; -
KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match 20.0%; Score 946.5; DB 16; Length 819;
Best Local Similarity 63.8%; Pred. No. 9.7e-38;
Matches 199; Conservative 23; Mismatches 41; Indels 49; Gaps 7;

QY 2 QIYTTDEIQVAKLAGKTYTDEGYIFD-----TSWIKKDSLSEAE 42
DB 511 QIYTTDEIQVAKLAGKTYTDEGYIFDPRDITSDGDAYVTPMTHSHWIKKDSLSEAE 570
QY 43 AAQAQAYAKEGLTPPSTDHQDSGNTAKGAEAIYNRVKAARKVPLDRMPYNLYQYTVEN 102
DB 571 AAQAQAYAKEGLTPPSTDHQDSGNTAKGAEAIYNRVKAARKVPLDRMPYNLYQYTVEN 630
QY 103 GSLLIPSYDHYHNKFWFDEGLYEAPKGYSLDGLATVYKYYE-----PR-----NAS 151
DB 631 GSLLIPHYDHYHNKFWFDEGLYEAPKGYSLDGLATVYKYYEHPNRPNSDNGFGNAS 690
QY 152 DHVRKNK-----ADQSKPDEKDEHSEVTPHPSDEKENHAGLN-----SADNL 198
DB 691 DHVRKNKGADNTNTEKPEKQTEKPEETPREKPEKPEKPEKPEKPEKPEKPEKPEK 750
QY 199 YKSTDTTEETEE---EADTTDEAIDGTPSIRONAMETTLGLKSSLLLTGKNNWISAE 255
DB 751 EEPQVETEKVEKLEAEADLLGKIQ----DPIKSNAKETTLGLKNNLLFGTDQNTIMAE 807
QY 256 VDSLALLKESQ 267
DB 808 AEKLLALLKESK 819

RESULT 4
Q9AG74 PRELIMINARY; PRT; 844 AA.
AC Q9AG74
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE PhpA.
GN PHpA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21246685; PubMed-11349048;
RA Zhang Y., Masi A.W., Barniak V., Mountzouras K., Hostetter M.K.,
RA Green B.A.;
RT "Recombinant PhpA Protein, a Unique Histidine Motif-Containing Protein
RT from Streptococcus pneumoniae, Protects Mice against Intranasal
RT Pneumococcal Challenge";
RL Infect. Immun. 69:3827-3836(2001).
DR EMBL; AF340221; AAK26629.1; -
SQ SEQUENCE 844 AA; 94769 MW; D738A55290FF8902 CRC64;

Query Match 20.0%; Score 946; DB 2; Length 844;
Best Local Similarity 62.5%; Pred. No. 1.1e-37;
Matches 202; Conservative 22; Mismatches 33; Indels 56; Gaps 8;

QY 2 QIYTTDEIQVAKLAGKTYTDEGYIFD-----TSWIKKDSLSEAE 42
DB 531 QIYTTDEIQVAKLAGKTYTDEGYIFDPRDITSDGDAYVTPMTHSHWIKKDSLSEAE 590
QY 43 AAQAQAYAKEGLTPPSTDHQDSGNTAKGAEAIYNRVKAARKVPLDRMPYNLYQYTVEN 102
DB 591 AAQAQAYAKEGLTPPSTDHQDSGNTAKGAEAIYNRVKAARKVPLDRMPYNLYQYTVEN 650
QY 103 GSLLIPSYDHYHNKFWFDEGLYEAPKGYSLDGLATVYKYYE-----PR-----NAS 151
DB 651 GSLLIPHYDHYHNKFWFDEGLYEAPKGYSLDGLATVYKYYEHPNRPNSDNGFGNAS 710
QY 152 DHVRKNK-----ADQSKPDEKDEHSEVTPHPSDEKENHAGLNPSADNLYKPTDTE 206
DB 711 DHVRKNKGADNTNTEKPEKQTEKPEETPREKPEKPEKPEKPEKPEKPEKPEKPEK 761
QY 207 ETEEA-EDTTDEAIP-----GTPSIRONAMETTLGLKSSLLLT 244
DB 762 ETEESPESPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPE 821
QY 245 GTRDNNTISAEVDSLALLKESQ 267
DB 822 GTQDNNTIMAEKLLALLKESK 844

RESULT 5
Q9ANY3 PRELIMINARY; PRT; 819 AA.
AC Q9ANY3
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Pneumococcal histidine triad protein B precursor (fragment).
GN PHTB.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21101045; PubMed-11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
DR EMBL; AF318954; AAK06759.1; -
KW Signal.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 819 819

```

JENCE 819 AA; 92108 MW; E602CFC16CC29A5F CRC64;
 Query Match 19.8%; Score 935.5; DB 2; Length 819;
 Best Local Similarity 63.1%; Pred. No. 3.3e-37;
 Matches 197; Conservative 23; Mismatches 43; Indels 49; Gaps

QY	2 QIYTTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLSEAPR 42
Db	511 QIYTTDEIQVAKLAGKYTAEDGYIFDPRDITSDGDAYTTPMHSHHWIKKDSLSEAPR 570
QY	43 AAQAYAKEGLTPPSTDHQDSGNTAKGAEAIIYNNVKAAKKVPIDRMPYNLOIYTVVKN 102
Db	571 AAQAYAXEGLTPPSTDHQDSGNTAKGAEAIIYNNVKAAKKVPIDRMPYNLOIYTVVKN 630
QY	103 GSIIIPSYDHYHNKIFWEPDEGLYEAAPKYSLEDLATYKYVVE-----PR-----NAS 151
Db	631 GSIIIPHYDHYHNKIFWEPDEGLYEAAPKYSLEDLATYKYVVEHPNERPHSDNGFGNAS 690
QY	152 DHYRKNK-----ADQSKPDEDEHDEVESEPHSPESDEKENHAGLNP-----SADNL 198
Db	691 DHYRNKNGQADTNOYEKPSSEKPOTEKPEETPREEKPOSEKPSKPTPEESPEES 750
QY	199 YKPSDTTEETEE---EAEQDTEAEIYGPSTPSTRONAKETLTGLKSLIGTKDNNTISAE 255
Db	751 EEPQVETEKVEELRAEDLLGKIQ---DPIIKSNAKETLTGLKNNLLFGTODNNTIMAE 807
QY	256 VDSLALLKRESQ 267
Db	808 AEKLLALLKESK 819

RESULT 6
 Q97QM8 PRELIMINARY; PRT; 802 AA.
 ID Q97QM8
 AC Q97QM8
 DT 01-OCT-2001 (TreeBLrel. 18, Created)
 DT 01-OCT-2001 (TreeBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TreeBLrel. 19, Last annotation update)
 DE Conserved domain protein.
 GN SP1175.
 GN Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RC MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae."
 RL Science 293:498-506(2001).
 RL EMBL; AE007418; AAK75284.1; -
 DR TIGR; SP1175; -
 KW Complete proteome.
 SQ SEQUENCE 802 AA; 90080 MW; 4FE5C8364EEA1833 CRC64;

Query Match 13.5%; Score 636; DB 16; Length 802;
 Best Local Similarity 45.3%; Pred. No. 8.5e-23;
 Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps

QY	2 QIYTTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLSEAPR 42
Db	513 QIYETEDEVRIQAQADKYTSDGYIFDEHDIISDGDAYTTPMHGSHHWIGKDSLSDEK 572
QY	43 AAQAYAKEGLTPPSTDHQDSGNTAKGAEAIIYNNVKAAKKVPIDRMPYNLOIYTVVKN 102

Db 807 PSSVSKEKI 815

RESULT 8
Q93GT5

AC Q93GT5 PRELIMINARY; PRT; 825 AA.
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE Histidine triad protein of group A streptococci.
 GN HTPA.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SSI-9;
 RA Terao Y., Kawabata S., Hamada S.;
 RT "Characterization of a novel histidine triad protein of group A
 streptococci";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB073859; BAB71774.1;
 SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFFB CRC64;

Query Match 4.9%; Score 233; DB 2; Length 825;

Best Local Similarity 19.1%; Pred. No. 0.0023;

Matches 182; Conservative 147; Mismatches 320; Indels 302; Gaps 49;

QY 65 GNTFKAAGAEIYNVKAARKVPLDRPYNLYQTVVKNGLSLIPSDVHYH-----114
 Db 47 GKAKAPKNTKMDQISAEEGISAE-----QIVVKITDQGYVTHGDNHFFYNGKVPYDA 100
 QY 115 -----NIKFWEFDEGLYAPKGYSLDGLATVYV--EPRNASDHYRKNK--A 159
 Db 101 IISEELLMTDPNRYFKQSDV--INEILDGVYK--VNGNYVYVYKPGSKRNIRTKQQA 156
 QY 160 DQSKPDEKDEHVESEPTPESEDEKENHAGLNPSADNLYKPTDTEEEEAEDTTDEA 219
 Db 157 EQVAKGTKEAKERGLAQVAHL--SKEEVAVN-----EAKRGQRYTDDG 199
 QY 220 EIPGTPSIRQNAMETLGLKSSLL-----LGTKNNTISAEDVSLALLKESQ--- 267
 Db 200 YIFSPTDI-----IDDLGDAYLVPHGNHYHVPKDLG--PSELAQAQVWSQKGRG 250
 QY 268 -----PAPIQG-----POIGQPTLPNNSLATPSPSLPINPCTSHEKH----- 304
 Db 251 ARPSDYRPTAPAPGRKAPIDVTPNPGQGHOPDNGGYHPAPRPND--ASQNKHORDE 308
 QY 305 -----EEDGYGFDAKRIIAEDSGFVMSGDSNHYFFKKDLT 341
 Db 309 FKGTFFKELLQLHRLDKYRNVHEEDGLIEPTQVIKSNAGFYVPHGDHHLIPRSQLS 368
 QY 342 EEQKAQAKHLEEVKTHSHGLDSSSEHQDYPGNAKEMDKLDKKIEKIAK--IMKQYGV-- 399
 Db 369 PLEMELADRY-----LAGOTEDDSSDSKPSDKVETHTFLGHRIKAYGKG 415
 QY 400 -----KRSEIVNKE-----KNAIYPSGDHHDADPIDEKPVGICHSHNYELFKP 446
 Db 416 LDGKPYDTSDAYVFSKESIHSDKSGVYAKRGHDFH-----IGFGL--EQYLEDV 466
 QY 447 EEGVAKKGNKVTYGEELTNVNLKKNSTFNQNTLANGOKRVSFSPFPELEKKLGIM 506
 Db 467 ANWY--KAGQ-----ADELAALD-----EQQEKERPLFD-----YKVS--- 500
 QY 507 LVRLITPDGRV--LEKYSKRGVGEVGNIANFELDQYLPFGQTF--KYTIASKQYPEVSYD 563
 Db 501 --RKVTGDKGVYMPKDGKY-----FYARDQLDLATQIAFAEQELMLKDKKHRYD 550
 QY 564 GTFTVTPSLATKMASQYIFPFHAGDYLYLRVNPQFAPVAPKGTDALRVFDEPHGNAYLENN 623
 Db 551 ---IVDTGIEPLADVSSSLPMHAGNATYDTGSSFVIPH-----IDHITH----- 591

QY 624 KYGCEIKLPKLNQGTTRAGNIKIPVTFMANAYLDNQST--YIVEVP-----ILEKENQ 676
 Db 592 -----VVPYSMLTR-----DQIATIKYVQHPVTRPDVWSRPGH 625
 QY 677 TDKPSILPQFKRNKAQENSKLDERVEEPK-----TSERKEKELSETGNSTSTLEVP 731
 Db 626 EESGSVIPNV-----TPLDKRAQENWQIHSAEVQK--ALAEGRFATPDGYI----- 672
 QY 732 TVDPVQEKVAFASGYMKLENVLFNMDGTIELYLPDSGEVIKK--NMADFTGEAPQNGEN 790
 Db 673 -FDP--RDVLAK--ETFVWK-----DGSFSPRADGSSLRITNKSDLSQAQWQAQEL 720
 QY 791 KPSENGKYSTGTVENQPTENKPADSLPE--APNKKVPKPNSTONGMLNPGNVGSDPML 848
 Db 721 LAKNAGATDT--DKPKEQKQAKSNENQPSSEAKSEESDDFI-----DSLP 769
 QY 849 DPALFEAPVDPVQEKLEKFTASYGLGLDS--VIFNMDGTIELRLPSEVI 897
 Db 770 DYGLDRATLEHINLAQK-----ANIDPKYLIFQPEG--VOFYKNGELV 813

RESULT 9

Q99XV4

AC Q99XV4 PRELIMINARY; PRT; 825 AA.
 DT 01-JUN-2001 (TREMELrel. 17, Created)
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE Hypothetical protein SPY2006.
 GN SPY2006.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE=21192684; Pubmed=11296296;
 RA Perretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006623; AAK34688.1;
 KW Hypothetical protein; Complete proteome.

Query Match 4.9%; Score 230; DB 16; Length 825;

Best Local Similarity 19.1%; Pred. No. 0.0032;

Matches 182; Conservative 146; Mismatches 321; Indels 302; Gaps 49;

QY 65 GNTFKAAGAEIYNVKAARKVPLDRPYNLYQTVVKNGLSLIPSDVHYH-----114
 Db 47 GKAKAPKNTKMDQISAEEGISAE-----QIVVKITDQGYVTHGDNHFFYNGKVPYDA 100
 QY 115 -----NIKFWEFDEGLYAPKGYSLDGLATVYV--EPRNASDHYRKNK--A 159
 Db 101 IISEELLMTDPNRYFKQSDV--INEILDGVYK--VNGNYVYVYKPGSKRNIRTKQQA 156
 QY 160 DQSKPDEKDEHVESEPTPESEDEKENHAGLNPSADNLYKPTDTEEEEAEDTTDEA 219
 Db 157 EQVAKGTKEAKERGLAQVAHL--SKEEVAVN-----EAKRGQRYTDDG 199
 QY 220 EIPGTPSIRQNAMETLGLKSSLL-----LGTKNNTISAEDVSLALLKESQ--- 267
 Db 200 YIFSPTDI-----IDDLGDAYLVPHGNHYHVPKDLG--PSELAQAQVWSQKGRG 250
 QY 268 -----PAPIQG-----POIGQPTLPNNSLATPSPSLPINPCTSHEKH----- 304
 Db 251 ARPSDYRPTAPAPGRKAPIDVTPNPGQGHOPDNGGYHPAPRPND--ASQNKHORDE 308

Query Match 4.78; Score 223.5; DB 2; Length 822;
Best Local Similarity 19.68; Pred. No. 0.0065;
Matches 186; Conservative 143; Mismatches 319; Indels 301; Gaps 52;

RESULT 11
Q8RYN2
ID Q8RYN2
PRELIMINARY;
PRT; 1524 AA.

RESULT 11

DE P0663E10.13 protein.
GN P0663E10.13
OS Oryza sativa (japonica cultivar-group). Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eubryotidae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RL clone: P0663E10."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF004317; BAB90730.1; NCBI_TaxID=39947;
SQ SEQUENCE 1524 AA; 168701 MW; E47FE1A9C5842228 CRC64;

Query Match 4.6%; Score 215.5; DB 10; Length 1524;
Best Local Similarity 18.9%; Pred. No. 0.036;
Matches 189; Conservative 139; Mismatches 371; Indels 299; Gaps 42;

QY 56 PSTDQDQSGNTEAKG-----AEIYNVKAACKVPLDMPYNLOY 96
DB 72 PPAARDGAGAGYGEIFRFDGFAEPYEDLLAEVALAAEIASSESSKS-----125
QY 97 TVEVKNSL-LIPS--YDHYHNKFP-EMFDEGLY-----EAPKGYSLDILLAT-----140
DB 126 SVKESGOLDADPSIIHQHYSYGVYDQHDEDEFSISPPDPSGKQFMSYNKATGRPD 185
QY 141 -----VKYVPEPRNASDRVRKNKADQDQSKPDEKHEDEVSEP-----THPES 182
DB 186 DIVKMTTCVPEPISYVDSRNIS-----NKSAMDQVVVVDCDTEFANGKSGMLFPSS 240
QY 183 DEKENHAGLNPADNLYKP-----STDTEETEEAEEDTDEAEIPTGTS-----IRQ 229
DB 241 SLSKASDSDVADQNLHTPICHPIKNDCEDEYHKLSTHSASSEVPSPDYPPFLRVSN 300
QY 230 NAKET-----LTGLKSLLLGTCKNNITIASVDSLLALKKESQAPATQGPQ 275
DB 301 NSLHTQPIKVPOLLAPSKLLNKESKANGKSTGLTFPSSS-----SVKASDPMADQN 357
QY 276 IGOPT-----LPNLSLAT---PSPSLPINPQTSHEKHEDEYGFDAHR 315
DB 358 LHPTCHPIKTCDEDRYHKLSTHSASSEDVSPDPYPLRVNPNLSLHTQPIKVPQPSK 417
QY 316 IIADESGFVMSGDSN-----HYFFKK--DLTEQIKAAQKHLKLEVTS-----358
DB 418 LLAKKES---KANGDSVSTNSAAAAAIAKEAMFEARLKAARLKAERKGDGSKLKKKP 474
QY 359 --HNGDLSLSSHEQDYPGNAKEMKDLKTEEKIAGIMQYGVKRESIVYVK--EKNAIL 414
DB 475 GHNGYKSTELKSMAP---EVRVYDEKL--TWRRIVKEEYETEAALYKNGDSSAVN 529
QY 415 YPGDHHADPIDHKKPVGIGHSHN-----YELFKPEGVAKKGNKRYTG 461
DB 530 LTHCDHNEKGVLPQRPQHTAQSGKLEQLGKWTSGAEFTVLISPDQ---KCKTNSV---583
QY 462 EELTNVYLLKNSTFNQNTLNGQKRVSFSPPELEKLGINMLVKLITPDGKVLKV 521
DB 584 -----TCGDNVOTNPSSKLG-----QPEKKG-----ETT 610
QY 522 SKRVFGGV---GNIANFELDQPLPGQTFKTYIASKD-----YPEVSYDGTFTVPT 570
DB 611 SGDVPVGGKWDGQDIAELRMEHVNLAR---EYALGSTEDEGCKAPTAPEISFSN--EKPT 664
QY 571 SLAYKMASQTIFFPHAGDYTLRVNPOFAPVKGTDALRVDFPHGNAYLNNYKVGGEIK 630
DB 665 ---YQESTETHFKECVGAQNT-----QERYGDGGAFTSCVDSK 701
QY 631 LPIPLNQGTTRTA-----GNKIP-----VTFMANAYLDNQSYIYVEVPILEKNG 676
DB 702 LHAPEIFCASLESCIGHCNGKNSPDSASTKETTSLGSKENKNNTEALEVPCAD-EMQ 760

QY 677 TDRPSILPQFKRNKAQEN-----SKLDEKVEBKPT-----SEKVEKEL- 715
DB 761 SOILQYHEPRNENIDKASQVKSLEESVEYETPNFQKSSSTAHTGETETVEKQMF 820
QY 716 ---SETGNSTNSILEEPTVTPQVEKAKFA--ESYGMKLENVLFNMDGTIE-----763
DB 821 SFSDELRPQNKIGNITEAPPESLIHKEIKFKGTETKAYITILEGDDVQVQKSGSLREANITL 880
QY 764 -----LYLPSGEVINKNMAFTGAPQNGENKPKSPENKSVSTGVENOPT 808
DB 881 ESASANEAEAEANAFVEGINVMEHYSY-STSVEDSDIQDSNRMDGMDGLVSHGN 939
QY 809 ENKPADSLPEAPNPKVPENSTDNGLNPNCGVSDP 846
DB 940 EBAAKDPWLDNSEKQVEEIFSHEEGQLSVGGIDGPP 977

RESULT 12
QY0459 PRELIMINARY; PRT; 5458 AA.
AC Q90459; O9GPG4;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE Erythrocyte membrane-associated giant protein antigen 332.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
RT "Molecular cloning and structure analysis of the Plasmodium falciparum
erythrocyte membrane-associated giant protein Ag332 (Pf332) gene."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF20180; AAF15293.3;
DR InterPro: IPR001313; Pumi110/Puf.
SQ SEQUENCE 5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;

Query Match 4.5%; Score 213.5; DB 5; Length 5458;
Best Local Similarity 19.2%; Pred. No. 0.26;
Matches 197; Conservative 181; Mismatches 419; Indels 227; Gaps 44;

QY 6 TDEIQVAK-LAGKTYTDEGYIFDTSWIKDLSLSE---AERAAQAYAKEGLTPPSTPH 61
DB 1114 SDEIVVEGSSATEIVVEGSSATEVV-VVEGSSVEEMLEEGSATEVVEEGL---SSD 1169
QY 62 QDSGNTGAEAGAIYNRVAAK-----KVPLDRMPY-----NLQYTVKNGS 104
DB 1170 VOKSGVIENVGSIYS-VTKAKDESMEKIPLEKSFVDDSEFKGGPTDNVSDVNSE 1228
QY 105 LIIPSYDHYHNKFEWDEGLVEAPKGYSLDILLAT-----VKYVPEPRNASDVRKKA 159
DB 1229 DIINEHTPLEETKIE-----ELPTEYTTADIHTRGETEYKLYKINEVEKAKF 1281
QY 160 QDSKPD-----EDKEHDEVSEPT-HPSDEKENAGLNPADNLYKPTDTEETEEEA 212
DB 1282 QEEKITENIPVERSEVTDIVQPSLAQVEQKE-----SDTNEIEE-- 1323
QY 213 EDTTDEAEIPGTSIQONAMETITGLKSLLLGTCKNNIT---ISAEDVSLALLKESOPA 269
DB 1324 TKLANEKIIPV-SVTENVVEK-EGLOTTEVLEEDSEITEEIVEEVSACEEIVEE--- 1379
QY 270 PIQGPQIGOPTLPNNSLATPSPSLPINPQTSHEKHEDEYGFDAHRIIADESGF--VMS 327
DB 1380 -----SSSEIVEE--ESSSEIVEEESSEIVEE 1409
QY 328 HGDSNHYFFKADTEPQIKAAQKHLKLEVTSNGLDLSLSEHODY-----PGNAKEMKL- 382
DB 1410 EESS---SEEIVEESSESSEIEEVSSTEVEEGSVTEEIVEEVSSTTEEVKDIG 1465
QY 383 ---DKKIEKIAK-----IMQYGVKRESIVNKEKNAILYPSGDH-----HHADPIDE-- 428


```
Query Match      4.4%; Score 208; DB 5; Length 2647;  
Best Local Similarity 19.9%; Pred. No. 0.18;  
Matches 201; Conservative 151; Mismatches 384; Indels 274; Gaps 48;  
  
QY   58 STDHQDGN-----TEAKGAEAYNNRVKAAK-----KVPLDRMPYNQLYTVE-- 99  
    | ||||| : ::::| :| :| :  
    : : : : :
```

01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Rhostry protein.
OS Plasmodium voelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=95021522; PubMed=7935623;
RA Keen J., Sinha K., Brown K., Holder A.;
RT "A gene coding for a high molecular mass rhotry protein of Plasmodium
RL Mol. Biochem. Parasitol. 65:171-177(1994).
DR EMBL; L27838; AAA21304.1;
SQ SEQUENCE 2269 AA; 265158 MW; F3D8CB103F9A6E1 CRC64;
Query Match 4.3%; Score 202.5; DB 5; Length 2269;
Best Local Similarity 19.4%; Pred. No. 0.26;
Matches 205; Conservative 158; Mismatches 344; Indels 351; Gaps 51;
QY 2 QITYTDEI-----QVAKLAGKYTTEDG-----YIFDTSWI 32
Db 76 ELTKSNEETVRLEKELFKYLLDEAEKRYLEGKLELNKKIKIDIIAKIEYVKNTEL 135
QY 33 KDSLSAEARAAQAVAKAGLTPPTDQSDGNTGAKAEALYNRVKA----- 81
Db 136 KK-----EIEK--NNAYIDELANSP---YKVTGYIENK--NTIYTKISYFOIYEGDID 184
QY 82 -----AKVPLDRMPNYLYTVEVKN--GSLIIPSYDHYHNINFEFDEGLYBAPKG 131
Db 185 TTFNELSSIVKEPID-----DIEDKYLENLRKSIDNVIDYKIQMEIETVKSHLNINETN 240
QY 132 YSLDLATVYVPEPNASDHVRK--NKADQSK-----PDEKHEDEVSEPH 179
Db 241 NKLPTITILEIKKY-----YDEISKELNKLMEDFKNEKELSKNSIDYDKKQQLSE-YK 294
QY 180 PESDEKENHAGLNPADNLKPPSTDEET--EERAEPTDDEAETPGTSPSIRONAMETLTGL 238
Db 295 SKMLEIRNH-----TNSQTNVDNTEEEAKQNYDASN-----EHW-- 330
QY 239 KSSLLGTCKDNNTISAEVDSILLALKESOPAPIQGPQIGPQITLNNSLATPSPSLPINPG 298
Db 331 -----TIPTNEDESKIISVK-----TKRDEILSKVNTYIDFN-- 364
QY 299 TSHKHEHDGVPANRIIADESGFVNSHGDSNHYFFKKDLTEQIKA--AKHLEEVK 356
Db 365 ---KKYKE-----TVNSEHSQFT-----ELT-DRIKAEVSDRELAKCE 398
QY 357 TSHGLDLSL-----SSHEODYPGNAKEMKOLD-----KKIEERIA----- 391
Db 399 QSPNDKSLINETHNINSIEKEYO--NINTLKVDEYIKVCKSTKESITKFSKQITLKDMLN 457
QY 392 -----GIMKOYGVKRESIVYVNEK---KNALIPSGDHHADP-----IDHKP 431
Db 458 QNKTVKETNSIDRSYIEKFEQILTGKTLKLENFTEFLNHNANNEELIKYFSDKAN 517
QY 432 VGIGHSHSNYELFPPEEGVAK--REGNRYTGEELTVNVLNKLKSTNNQNTLANGOKR 489
Db 518 LGINEENLYNQFTEKTFNDNIEKN--IHNEELSKIEIKHASYN----- 564
QY 490 VSPFPPPELEKLGIMLVKLITPDGKLVKSGVKGEGVGNIANFELDQYLPQGTFFK 549
Db 565 -----ISETEREIGIN-----IESLNTKVPEKVENVTNL-----NKIK 599
QY 550 YTIASKDPEVSYDGTFTVPSLAYKMASQITFFPHAGDYLRVNPQ--FAPVKGTDALV 608
Db 600 EKLKHYDFDFGKEG-----NIKYT-----DKIKINDIMAVSQOIQHI 640
QY 609 RVFDEPHGNAYLENKYKEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIYEV 668
Db 641 NGLDDIQKKS-----EST--VSEMEKEQINKLEK-----VSTNEISDN-----V 677
QY 669 PILEKENQ-----TDRPSILPOFRKKAQENSKL-----DERVEEPKTS--ERKEKEKLSETG 719

Db 678 EGKKKKQIIVTKIDK-----KKNIYEINKLLSEIKERDNTSLEKVKDINLS-YG 729
QY 720 NSTSNSTLEEVPTVDPVQEKVAKFASGYMKLENVFNMDGTIELYLPSEGVTKKNADF 779
Db 730 QNLGNLFLEQIDEEKKAENTIKSMEAYIDDLDNK--KKSQEIEIETMDIKMDINKKEAL 788
QY 780 TGEAPOGNGENKPSKNGKVGTVENOPT-----ENKPADSLPEAPNEKP 824
Db 789 KISHDDKKCHDSKNNIKENISDIYDKSSLIQDFSRESINDIKKNLQKNVSESN--- 845
QY 825 VKPENSTDNGLMPEGNVGSDPMDLPALEBAPVDPVQEKLEKFTASYGLGL---DSVIP 881
Db 846 ---HNSDINQCLNEVANIYNILKIKKIIDKVKYETSEIKENKKNINDELANSSEKVIK 902
QY 882 NMDGTIELR-----LPSGEVTKKNL 901
Db 903 KIEGDLSLKECRSKINSTLDDKDIDECIKNINVLKNI 940
RESULT 15
Q963T1 PRELIMINARY; PRT; 1078 AA.
AC Q963T1;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Glutamate-rich protein (Fragment).
GN GLURP.
OS Plasmodium reichenowii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21313688; PubMed=11420113;
RA Thelsen M., Thomas A.W., Jepsen S.;
RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
RT glutamate-rich protein (GLURP) from Plasmodium reichenowii."
RL Mol. Biochem. Parasitol. 115:269-273(2001).
DR EMBL; AF356828; AAK40236.1;
FT NON_TER 1 1078
SQ SEQUENCE 1078 AA; 123581 MW; 8D388D88B223913C CRC64;
Query Match 4.3%; Score 201.5; DB 5; Length 1078;
Best Local Similarity 19.9%; Pred. No. 0.11;
Matches 178; Conservative 152; Mismatches 357; Indels 209; Gaps 42;
QY 11 QVAKLAGKYTTEDGYFDTSWIKKDSLSAEARAAQAVAKAGLTPPTDQSDGNTGAK 70
Db 312 EIFSLDKETTNEDILPNQNPENLKQSEINHVQDHALPRENIDIKLDNKEHIDESQ 371
QY 71 GAELIY--NRVAAKVPDRMPYNYLYTVEVKN--GSLIIPSYDHYHNINKEFWFDEGLY 126
Db 372 HNVNVLQENNNHNLQLEPQEK--NIE-SFEPKNTDSEILPENVEKEIYVD-----V 422
QY 127 EAPKGYSLDLATVYVPEPNASDHVRKNKADQSDKDEKHEDE--VSETHPESDEK 185
Db 423 PSPKPHNHETL-----EQETSESEH---EEAVSEKNAHETVEHEEAVSQSNPEKADN 472
QY 186 ENHAGLNPADNLKYPSTDETEEEAEEDTDEAETPGTSPSIRONAMETLTGLKSLILG 245
Db 473 DGNVSQNSNNELENFVSEKSEHEPAENEE-----SLEEAHQEIV----- 516
QY 246 TKNNTISAE---VDSLALLKESOPAPIQGPQIGPQITLNNSLATPSPSLPINPTSHE 302
Db 517 PEQNNQSGESKLVDNDEGGFEAHQBEIYVEQ-----NNQESGESKLVDNDEGGFEE 569
QY 303 KHEED-----GYGFANRIIADESGFVNSHGDS-----NHYFFKDLI---- 341
Db 570 ARQBEIYVEPQNNQSGESKLVDNDEGGFEAHQBEIYVEQ-----VSTNEISDN-----V 677
QY 342 -----EQIKAAQKHLVEEYKTSHTNGLDLSSSHEDYPGNAKEMKOLDKKIEEKIAG 392

Search completed: March 27, 2003, 11:56:59
Job time : 118 secs

Fri Mar 28 09:21:10 2003

us-09-884-465a-332.rapb

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:57:08 ; Search time 20 Seconds
(without alignments)
2660.182 Million cell updates/sec

Title: US-09-884-465A-332
Perfect score: 4728
Sequence: 1 MQITYTDEIQVAKLAKYT.....IELRLPSGEVKKNLSDFFA 906

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues
Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	818.5	17.3	763	10	US-09-765-272-66 Sequence 66, Appl
2	636	13.5	796	10	US-09-765-272-56 Sequence 56, Appl
3	431	9.1	447	10	US-09-765-272-182 Sequence 182, Appl
4	226.5	4.8	793	9	US-09-252-088-15 Sequence 15, Appl
5	216	4.6	715	9	US-09-252-088-16 Sequence 16, Appl
6	197	4.2	1478	10	US-09-801-368-52 Sequence 52, Appl
7	192	4.1	571	10	US-09-765-272-4 Sequence 4, Appl
8	190	4.0	1596	9	US-09-902-432-4 Sequence 15, Appl
9	181.5	3.8	1139	9	US-09-820-843A-15 Sequence 45, Appl
10	180.5	3.8	1183	9	US-09-870-759-45 Sequence 5635, Ap
11	180.5	3.8	2368	10	US-09-815-242-5635 Sequence 12389, A
12	180.5	3.8	2368	10	US-09-815-242-12389 Sequence 5816, Ap
13	178	3.8	2478	10	US-09-815-242-5816 Sequence 12967, A
14	178	3.8	2478	10	US-09-815-242-12967 Sequence 9, Appl
15	174	3.7	639	10	US-09-825-144-9 Sequence 4, Appl
16	170.5	3.6	1545	9	US-09-839-996-4 Sequence 6, Appl
17	170	3.6	1848	9	US-09-839-996-6 Sequence 5, Appl
18	169	3.6	639	10	US-09-823-240-5 Sequence 5456, Ap
19	168.5	3.6	978	10	US-09-815-242-5456

20	168.5	3.6	1001	10	US-09-815-242-12686 Sequence 12686, A
21	166.5	3.5	1113	10	US-09-815-242-5836 Sequence 5836, Ap
22	166	3.5	1616	9	US-09-820-843A-16 Sequence 8, Appl
23	166	3.5	2785	10	US-09-801-574-8 Sequence 23, Appl
24	165.5	3.5	1181	10	US-09-870-122-23 Sequence 5898, Ap
25	164.5	3.5	1349	10	US-09-815-242-5898 Sequence 13137, A
26	164.5	3.5	1349	10	US-09-815-242-13137 Sequence 1, Appl
27	164	3.5	1164	10	US-09-870-122-1 Sequence 3, Appl
28	163.5	3.5	1881	9	US-09-998-425-3 Sequence 316, App
29	163.5	3.5	1881	9	US-09-997-977-3 Sequence 911, App
30	162.5	3.4	766	10	US-09-801-368-316 Sequence 911, App
31	162.5	3.4	1242	9	US-09-925-299-911 Sequence 3, Appl
32	162.5	3.4	1242	10	US-09-925-299-911 Sequence 2, Appl
33	162.5	3.4	1781	10	US-09-738-877-3 Sequence 5845, Ap
34	162.5	3.4	26926	9	US-09-759-508B-2 Sequence 9, Appl
35	162	3.4	1016	10	US-09-815-242-5845 Sequence 2, Appl
36	162	3.4	2122	9	US-09-813-214A-9 Sequence 4, Appl
37	161	3.4	1167	10	US-09-870-122-2 Sequence 3, Appl
38	160	3.4	690	10	US-09-785-770A-4 Sequence 12955, A
39	160	3.4	714	10	US-09-785-770A-3 Sequence 5834, Ap
40	160	3.4	1111	10	US-09-815-242-12955 Sequence 12996, A
41	159	3.4	2437	10	US-09-815-242-5834 Sequence 114, App
42	159	3.4	6281	10	US-09-815-242-12996 Sequence 4, Appl
43	156.5	3.3	1125	9	US-09-974-298-114 Sequence 35, Appl
44	156	3.3	1332	10	US-09-982-091A-4
45	156	3.3	1597	9	US-09-832-292-35

ALIGNMENTS

RESULT 1
US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:

us-09-884-465a-332.rapb

21:10 2003

66

.tch 17.3%; Score 818.5; DB 10; Length 763;
 .ocal Similarity 66.1%; Pred. No. 3.2e-42;
 .nes 168; Conservative 16; Mismatches 29; Indels 41; Gaps 5;
 QY 2 QITTDDEIOVAKLAGYTTEDGYIFD-----TSWIKKDSLEAER 42
 DB 498 QITTDDEIOVAKLAGYTTEDGYIFDPRDITSDGDAYVTPMTHSHWIKKDSLEAER 557
 QY 43 AAAQAYAKEKGLTPPSTHDQSGNTEAKGAEAIYNNRKAAYKVPDLDRMPYNLOYTVEVKN 102
 DB 558 AAAQAYAKEKGLTPPSTHDQSGNTEAKGAEAIYNNRKAAYKVPDLDRMPYNLOYTVEVKN 617
 QY 103 GSLLIPSDHYHNIFEWFDGLYEAPKGYSLDILLATVYKVE-----PR-----NAS 151
 DB 618 GSLLIPSDHYHNIFEWFDGLYEAPKGYSLDILLATVYKVEHPNRPDSHNGFGNAS 677
 QY 152 DHVRKNK-----ADQSKPDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSDTDE 206
 DB 678 DHVRKNKQADNTNTEKPSKQTEKPEETPREKPSKQTEP-----KETEPE 731
 QY 207 ETEEAEDTDEAE 220
 DB 732 ESPESEEPQVETE 745

RESULT 2
 US-09-765-272-56
 ; Sequence 56, Application US/09765272
 ; Patent No. US20020061945A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 56:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 796 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
 ;
 ; US-09-765-272-56
 Query Match 13.5%; Score 636; DB 10; Length 796;
 Best Local Similarity 45.3%; Pred. No. 4.6e-31;

Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps 8;
 QY 2 QITTDDEIOVAKLAGYTTEDGYIFD-----TSWIKKDSLEAER 42
 DB 507 QIETDEVRIAQLADRYTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIKKDSLEAER 566
 QY 43 AAAQAYAKEKGLTPPSTHDQSGNTEAKGAEAIYNNRKAAYKVPDLDRMPYNLOYTVEVKN 102
 DB 567 VAAQAYAKEKGLTPPSTHDQSGNTEAKGAEAIYNNRKAAYKVPDLDRMPYNLOYTVEVKN 626
 QY 103 GSLLIPSDHYHNIFEWFDGLYEAPKGYSLDILLATVYKVE-----PR-----NAS 151
 DB 627 GNLLIPKDHYNIFEWFDGLYEAPKGYSLDILLATVYKVEHPNRPDSHNGFGNAS 686
 QY 152 DHVRKNKADQSKPDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSDTDETEEE 211
 DB 687 EHV-LGKKDHSEDPNKNFADE--EVEETPAEPE-----VPOVETEKVDAQ 730
 QY 212 AEDTDEAEI-----PCTPSIRONAMETLGLKSSLLGLTKDNNTISAEVDSLLALLKESQ 267
 DB 731 LK-----EAEVLLAKVTDSSLLKANATETLAGLRNLTQIMDNNSINAEKLLALLKGSN 786
 QY 268 PAPIQGPQI 276
 DB 787 PSSVSKEKI 795
 RESULT 3
 US-09-765-272-182
 ; Sequence 182, Application US/09765272
 ; Patent No. US20020061945A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
 ;
 ; US-09-765-272-182
 Query Match 9.1%; Score 431; DB 10; Length 447;
 Best Local Similarity 32.2%; Pred. No. 7e-19;

```
Matches 123; Conservative 40; Mismatches 91; Indels 128; Gaps 12;
QY 7 DDEIOVAKLAGYTTEDGYIF-----DTS-----WIKDLSSEAEAAAQA 47
Db 144 NSNAVARSQGYTTNDGYVNPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAAGA 203
QY 48 YAKEKGLTP-----PSTHQDSGNTAEGA-----EAINRVKAAKVPPL 87
Db 204 HLAGNMOPQSLSYSTASDNTQSVAGSTSKPANKSENLSQLKELYDPSAQRYSSES 263
QY 88 DRWPYNLQYTVVKNGLSILIPSDYDHVNIKPEWFEGLYEAPKGYSLDGLLATVYIYVEP 147
Db 264 DGLVDPDAKIIISRTNGVAIPGHDHYHPIYS-----KLSALEKIAM--VPI 310
QY 148 RNASDHVRNKADQSPDEKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPSDTDEE 207
Db 311 SGTGSTVSTN-----AKPNEVV-----SSLSGLSSN--PSSLTTS 343
QY 208 TEEAEEDTDEAEIPGTPSIRQNMETLTGLKSSLLTGKONNTISAEDVSLALLKESQ 267
Db 344 KE-----LSSASDGYIFNPKD-----IVEETA 365
QY 268 PAPI-----QGPOIGOPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDANRI 316
Db 366 TAVIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDANRI 425
QY 317 IAEDESGFVMSHGDSNHYFFKK 338
Db 426 IAEDESGFVMSHGDNHNYFFKK 447

RESULT 4
US-09-252-088-15
; Sequence 15, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clement
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; CURRENT APPLICATION NUMBER: US/09/252,088
; EARLIER FILING DATE: 1999-02-18
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 793
; TYPE: PRT
; ORGANISM: group B streptococcus
US-09-252-088-15

Query Match 4.8%; Score 226.5; DB 9; Length 793;
Best Local Similarity 19.08; Pred. No. 4.5e-06;
Matches 181; Conservative 148; Mismatches 316; Indels 307; Gaps 51;

QY 65 GNTAEKGAELYNRVKAAKVPIDRMPYINQYTYVEKNGSLIIPSYDHYH-----114
Db 18 GKARAPTKNTKDQISAEIGISAE-----QIVKITDQGVTSHGHDHYHYFNGKVPYDA 71
QY 115 -----NIFKFWDEGLYAPKGYSLDGLLATVYIYV--EPRNASHVRKKN--A 159
Db 72 IISELLATDPNRYFKQSDV--INEILDGYIK-----VNGNTYVILKPGSKRKNITROOIA 127
QY 160 DQSKPDEKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPSDTDETEEAEDTDEA 219
Db 128 EOVANGTKEAKGLAQVAHL-----SKEVAAYN-----EAKRQGYTTDDG 170
QY 220 BIPGTPSIRQNMETLTGLKSSLL-----LGTKNNTYSAEVDLSLALLKESQ---267
```

```
Db 171 YIISPDI-----IDDLGDALVPHGNHYHYIPKDKLS--PSELAQAAYWSQKQGRG 221
QY 268 -----PAPIQG-----PQIGOPTLPNNSLATPSPSLPINPGTSHKHEH-----304
Db 222 ARPSDYRPTPAPRRKAPIDVTPNPGQGHQPDNGGYPHPAPPND--ASQNKHQRDEPK 279
QY 305 -----EEDGYGFDANRIIAEDSGFVMSHGDSNHYFFKKDLTEE 343
Db 280 GKTFKELLDQHLRLDLKYRVEEDGLIFETQVIKSNAGFYVPHGDHYHIIIPRSQSLP 339
QY 344 QIKAAQKHLEEVKTSNGLDLSLSDSHEDQYPCNAKEMKDKLKKIEEKIAG--IMQYGV---399
Db 340 EMELADRY-----LAGQTEDNDSGSEHSPKSDKEVTHFLGHRKAIKAYGKGLD 386
QY 400 -----KRESIVVKE-----KNAILYPSGDHHDADIDEHKVPKIGHSHSNYELFPKE 448
Db 387 GKPYDTDAYVYFSKESHSVDKSGVTAKHGDHFIY-----IGFGEI--EQYELDEVAN 437
QY 449 GVAKEGKNVYTGSELTNVVNLKSNFTNQNFTLANGOKRVSFSPFPELEKKLGINMLV 508
Db 438 WV-TAKGQ-----ADELAALD-----QEQGKEKPLFD-----TKVYS-----469
QY 509 KLITPDGKV--LEKVSQKVGEGVGNITANFELDQYPLPGQTF-KYTTASKDYPEVSYDGT 565
Db 470 RKVTGKQKVGVMKPGDKDY-----FYARDQLDLTQIAFAEQELMKDKKHRYRD--519
QY 566 FTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYK 625
Db 520 -IVDTGIEPLAVDVSSLPHMAGNATYDTGSSFVIPH-----IDHIIH-----560
QY 626 VGEIKLPIPKLNQGTTRTAGNKPIPVTFMANAYLDNQST--YIVEVP-----ILKENQTD 678
Db 561 -----VVPYSWLTR--DQIATVKYVMQHPEVRPDVMSKPGHEE 596
QY 679 KPSILPQFKRNKAQNSKLDKVEBPK-----TSEKVEKEKLSGTGNSSTSLLEEPTV 733
Db 597 SGSVIPNV-----TFLDKRAGPNWQIISAEVOK--ALASGRFATPGYI-----P 642
QY 734 DPVQEKVAKFAESYGMKLENVFNMDGTIELYLPSEGEVIK--NMADTGEAPQNGENKP 792
Db 643 DP-RDVLAK--ETFWK-----DGSFSIPRAGDSSLRITINKSDLS----QAEWQQAQ 687
QY 793 SENGKYSTG--TVENOPTENPADSLPAPNEKP---VKPENSTDNGMLNPEGVNGSDPM 847
Db 688 ELLAKNTGTDATDTDKPEKQQAOKSNE--NQOPSEASKEEKESDDFI-----DSL 736
QY 848 LPALEAPAVDPVQEKLEKFTASYGLGLDS--VIFNMDGTIELRLPSEGEVI 897
Db 737 PDYGLDRATLEDHINQLAQK-----ANIDPKYLIFQPEG-VQFYKNGELV 781

RESULT 5
US-09-252-088-16
; Sequence 16, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clement
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; EARLIER FILING DATE: 1999-02-18
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 715
```

us-09-884-465a-332.rapb

21:10 2003

SM: group B streptococcus
J-088-16

dry Match 4.6%; Score 216; DB 9; Length 715;
Best Local Similarity 17.7%; Pred. No. 1.7a-05;
Matches 171; Conservative 127; Mismatches 271; Indels 396; Gaps 46;

QY 2 QITVDBE-----IQVAKLAKYTTEDGVIYDTSWIKKDSLSERAAQAAYAKEGLTTP 57
DB 66 QVHLSKEEVAANRQGRYTTDDGIIS-----P 97
QY 58 STDHDSNTEAKGAEAYINRKAAYVPLDRMPYNLOYVEVKNGLSLIIPSYDHYHNK 117
DB 98 DTIIDLDG-----AYLVPNGHYH--- 117
QY 118 FEWDEGLYEAAPKGYSLDILLATVYVYEPNASHVKNKADQDQSKDPDEKHEVSEP 177
DB 118 -----YIPKDLSPSELAAQAAY-----SQQGRGAREPSD----- 148
QY 178 THPSEDEKENHAGLPSADNLKYPSTDETEEEAEEDTDEAIPGTPSIRQNAWELTG 237
DB 149 -----YRPT-----PAPG----- 156
QY 238 LKSSLLGTKNNTISAEVDSLALLESQAPIQGPQIGQPTLNNSLATPSPSLPIMP 297
DB 157 -----RRKAPIDVTPNPGQHQPDNGGYHAPPPND- 189
QY 298 GTSHEKH-----EEDGYGFDAIRIIAEDSGFVNSHGD 330
DB 190 -ASQKHQDEFKTKFELLQHLRLDLKYRHYVEDGLIFETQVIKSNAGYVYPHGD 248
QY 331 SNHYFFKDLTEQIKAAQKHLEVTSHNGLSLSHEDQYPCGNKAKEMKDLKKEKI 390
DB 249 HYHIPSQSPLEMLADRY-----LAGQEDNDGSGSEHSPSKREVTHF 295
QY 391 AG-IMKOYGV-----KRESIVNKE-----KNALYPSGDHHAIDPEHKPVIG 435
DB 296 LGHRKAYGLDKPYDTSDAYFSEHSISVDKSGYAKHGDFHY-----IGFG 347
QY 436 HSHSYELFKPEGVAKGKVTGELTNVNLKNSTNNQNTLQNGKQVSYFPP 495
DB 348 EL-EQYELDEVANWV-KAQO-----ADELAAALD-----OEQKKEPLFD-- 386
QY 496 PELEKKGIMLVKLITPDGV--LEKVSQGVFGVGNIAFELDQPLPCQTF-KYTI 552
DB 387 ---TKVS-----RKVTGDKVGYMMPKDGKY-----FYARDQLDLQIAFAEQEL 430
QY 553 ASKDYPEVSDGTFVPTSLAYEMASQTFYFPHAGDTLYRYNPQFVAPKGTDALVRVD 612
DB 431 MLKDKKHYRD-----IVDTGIEPLAVDVSSLPWAGNATYDTSFVIPH-----ID 479
QY 613 EPHGNAYLENNYKVEIKLPIPLANOGTTRTAGNKIPVTFFMANAYLDNST--YIVVP- 669
DB 480 HIB-----VVPISWLR-----DQIATKVIYVQHPE 505
QY 670 -----ILEKENQTKPSILPOFKKKAQENSKLDEKVEEPK-----TSEKVEKLESETGN 720
DB 506 VRPDWSPKPGHESSGVPNV-----TPLDKRAGMPWQIHSAREVQK-ALAEGRF 556
QY 721 STNSSTLEEVTPVQEKVAKFAEYSGMKLENLYFNMDGTIELYPSGVEIHK-NWADF 779
DB 557 ATPDCYI-----FOP-RDVLAK--ETFWK-----DGSFSPRADGSSLRITNKS 600
QY 780 TGEAPOGNGENKPSNGKSTVG--TVNQPTENKPAISLEAPNEKP--VKPENSTNG 834
DB 601 S-----QAEWQQAQELLAKKNTGDATDTDKPREKQQAQKSNB--NOQPSAESEKESDDF 654
QY 835 MLNPGNVGSDPMLPALPEAPVDPVQEKLEKFTASTYGLGLDS--VIFNMDGTIELRLP 892
DB 655 I-----DSLPGYGLDRATLEHINOLAQK-----ANIDPKYLFQPEG-VQFYNK 698
QY 893 SGEVI 897

Db 699 NGELV 703

RESULT 6

US-09-801-368-52

Sequence 52, Application US/09801368

Patent No. US20020128250A1

GENERAL INFORMATION:

APPLICANT: Busby, Robert

APPLICANT: Cali, Brian

APPLICANT: Hecht, Peter

APPLICANT: Holtzman, Doug

APPLICANT: Madden, Kevin

APPLICANT: Maxon, Mary

APPLICANT: Milne, Todd

APPLICANT: Royer, John

APPLICANT: Salama, Sofie

APPLICANT: Sherman, Amir

APPLICANT: Silva, Jeff

APPLICANT: Summers, Eric

TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

FILE REFERENCE: 109272.147

CURRENT APPLICATION NUMBER: US/09/801,368

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: US 09/487,558

PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: US 60/160,587

PRIOR FILING DATE: 1999-10-20

NUMBER OF SEQ ID NOS: 440

SOFTWARE: Patent in version 3.0

SEQ ID NO 52

LENGTH: 1478

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-801-368-52

Query Match 4.2%; Score 197; DB 10; Length 1478;

Best Local Similarity 17.7%; Pred. No. 0.00064;

Matches 193; Conservative 139; Mismatches 307; Indels 404; Gaps 47;

QY 34 KDSLSERAAQAAYAKEKGLPPSTD-----HODSGNTEAKGAEAYINRV-----KAA 82

DB 251 QEDISNRSSTSEASLPSTKS-GPSKTDENFLHSTHQKTSASSLYRRSPTSLGSSS 309

QY 83 KVPIDRMPYNLOYTVBKNGLI-----IPSYD-----HYHNKKEP----- 119

DB 310 SNASSAKSPNKLISIPARPHSIIESNSTLTKSASPPSPSPSIFRRHHKSSSSSLL 369

QY 120 --WFDEGLYEA-----PKGYSLDILLATVYVYEPNASHVKNKADQDQSKDPDEK 171

DB 370 NSLFGSGIGEEAPTKPNPQGHSL-----SSENLAQK-----SKHYETNVS 410

QY 172 DEVSEPTIPEDEKEN--HAGLNPSADNLKYPSTDETEEEAEEDTDEAIPGTPSIQ 229

DB 411 SPLKQSSLTSDGKLNWKNPKKRSQIGVSPNTVAVYSQ-----TPSLKS 458

QY 230 NA-----METL-----TGLKSSLLIGTKONNTISAEVDSLALLKESQ 267

DB 459 NSSTATLVQADVNIIPSSSSPPPIPKTANRSLVISTEDTPKIS-----STTASFKETY 514

QY 268 PAPIQGPQIGOPTLNNSLATPSPSLPINCPTSHEKH-----BEDGYGFDAIR-- 315

DB 515 PDCINPDK-----TVP-----VPMNNQYKSVKFNLLDQKFFPLKKTGLNDSENKYI 560

QY 316 IIAEDSGGV-----MSH-----GDSNHYF 335

DB 561 LVTKDNVSVFVPLNKLKSVAKLSFKESALFKLGINHKNVTFHMTDFDCDICAIPDDTLEF 620

QY 336 FKDDL-----TEEQIKAAQKHLEEVKTSING--LDSLSHEQDYPCGNKAKEMKDL- 383

DB 621 LKKSFLNTSGKIYINDOMKLOQKPKAPLTSENNVPKSVKSKSMRSGTSSLIASITDD 680

384 -----KKIEKIAGIMKQY-----GVKRESIVVVKK 410
 Db VSIYSSDITSPDEHAGSGRRYPQPSYYIDRVNTNPTTELNYWNIKE---VLSHEE 737
 411 NA--IYPSGDHHDPIDEHK-----PVGIGHSHSNYELFKPEEVA---KEGKNVY 459
 Db NAKMVKFTSPKLEINLPDRGSKLINPTIPITENESKSFQVLRKDEGTEIDFNHRRSPY 797
 460 TGEEL-----TNVNLKNSFNQNTLANGQKRVFS-----FPPELEKKL 502
 Db TKPELAPREAPKPPANTSPORTLSTSK-QNKPILYRASTKISRSKSKPLPOL---L 853
 503 GIMLVKLITPDGKLVKSVKVFGEVGNITANFELDQPYLPFGOTFKYTIASKDYP----- 558
 Db 854 SSIPEASSSPD-----SLTSSYTPAST--HVLIPQYTKGAND 889
 559 -----EVSVDGTFYPTSLAYKM-----ASQTFYFPFHAGDTYLRVNPQFAVPGK 603
 Db VMRLKTDQDSTSP-SLKMVKQKVRNSVSTNSIFIS----- 929
 604 TDALVRVDFEFGHAYLNNYKVEIKLPIPKLNGTTRTAGNKIPVTFMANAYLDNOST 663
 Db 930 -----PSLLKRGSKRV---VSSTSAADIPENDIT 958
 664 YIVEVPILEKENQDKPSLPFOFKANKAQAENSKLDEKVEEPTSEKVEKELSETGNSTS 723
 Db 959 FADAPPFSDSDSDSS-----SSDDIINSKKKTAPETNNENKDKESONSSTHS 1009
 724 NSTLEVPTVDVQEVKVAFAESYGMKLENVLFNMDGTIELYLPSEGVKKKNMADFTGEA 783
 Db 1010 DEIFYDSQODKAKRWK-----FRPPEVYQNLKF---F 1043
 784 PQONGENKPSGKSVGTGVNQPTENKPADSL-----PEAPNEKPVKPNST 831
 Db 1044 PRAN-LDKPITTEGAS-----PTSPKSLDLSLSPKNVASSRTEPSTPS-RPVPPDSY 1094
 832 D-----NGMLNP 838
 Db 1095 EFTQDGLNGKNKP 1107

RESULT 7

US-09-765-272-4
 ; Sequence 4, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 571 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-765-272-4

Query Match 4.1%; Score 192; DB 10; Length 571;

Best Local Similarity 22.2%; Pred. No. 0.00037;
Matches 156; Conservative 88; Mismatches 197; Indels 262; Gaps 40;

QY 284 NSLATSPSLPINFPGTSHRHEEDGYGFDA-----RIIADESGFVMHSGHNSHYFFK 337
 Db 45 SSVATPT-----KQKVDYNNVTFVDHPVQVQAIQEQTPVSTKTEVQVVE 92
 QY 338 KDLTEEQIKAAQHLEVKTSNGLDLSLSEHQDYPGNAKEMKDLKKIEKTIAGIMKQY 397
 Db 93 KPFSTELINPR---KEEKSSDSQEQLAEH-----KNLETKKEKIS----- 131
 QY 398 GVKRESIVVKNKNAIYPSGDHHDADPIDE-----HKPVGIGHSHSNTELEKPEGVA 451
 Db 132 -----PKEKTV-----NLAPODEVLSGOLNKP-----ELLYREETME 165
 QY 452 KKEGNKYVTGELNPNVNLKNSFNQNTLANGQKRVFSFPPELEKLGINK-LVKL 510
 Db 166 ---TKIDFQBEI-----QENPDLAEGTVRV-----KQEGKLRKKEIVRI 202
 QY 511 IT-----PDGKLVKVSGR--VFGE-----GV-----GNIANFELDQ 540
 Db 203 FSVNKEEVSRIVSTTAPSPRIVEKGTQTKQVKEQPETGVEHKDVOSGAIIVEPAI-Q 261
 QY 541 PVLPGQTEKTYTIASKDPEVSYDGTFTVPTSLAYKMASQTIIFYPHAGDTYLRVNPQ--- 597
 Db 262 PQLP-----EAVVSDKGEVQVQ---TLPEAVV-----TDKGET--EVQPESPD 300
 QY 598 -FAPVKGTDALVRVDFEFGHAYLNNYKVEIK--LPKPL-NOGTTTAGNKIPVTEM 653
 Db 301 TVVSDKGEPEQVAPLPEYKGN-----IEQVPETPVETKKEGQPEKT--EEVPV--- 347
 QY 654 ANAYLDNQSTYIYVPILEKENQTDKPSILPQFKRN---KAQENSKLDERKVEEPTSEK- 709
 Db 348 -----KPEETPVNPNBEGTTEGTSI--QEAENPVQPAEESTTNEKV-SPDTSKN 395
 QY 710 -----VEKEKLSGTGNSSTNSLSEVFTVDVPOEVKVAFAESYGMKLE 752
 Db 396 TGEVSSNPSDSTTSVGSNKKPHNDKNSSEKTEVEVP-VNP----- 437
 QY 753 NVLFNMDGTIELYLPSEGVKKNMADFTGEAPQNGENKPSNGKV---STGTVENOPT 809
 Db 438 -----NEGTV-----GTSNQETEKVPQAEETQNSGKIANENGEVSNKPSD 481
 QY 810 NKP---ADSLPEAPNEKPVKPNKNS-----TDNGMLNPE-----GNVGSNDPMLDPALEAPVDP 860
 Db 482 SAPPVEESNQPEKNGTAKPKNSGNTTSENGQTEPEPSNGNSTEDVSTESNTSNGNEE 541
 QY 861 VOEKKLFTASYGLGLDSVIFNMDGTIELRLPSGVEIKKNSLD 903
 Db 542 IKQENE-----LQPDKKEVEPEKTELELR-----NVSD 568

RESULT 8

US-09-902-432-4
 ; Sequence 4, Application US/09902432
 ; Patent No. US20020160002A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Irwin H. Gelman
 ; APPLICANT: Susan G. Jaken

[illegible]

Query Match 3.8%; Score 181.5; DB 9; Length 1139;
Best Local Similarity 19.7%; Pred. No. 0.004;
Matches 161; Conservative 118; Mismatches 342; Indels 197; Gaps 39;

[illegible]

[illegible]

41:10 2003

666 VEVPILEKENOTDKPSILPQFNRKNAQNSKLDKVEPKTS-----EKVEKEKL 715
 1042 TDVNAKTNEATIAAITPD-ANVPTAKQAIADVQAOQATDAIDANNGATTEKAAKQO 1100
 716 SETGNSTNSILEEVTVDVQCEKAPAEYSYGMK-LENVFNMDGTIELYLPSEVYK 774
 1101 VQTEKTADTAIDGAHTNAEVE--AAKNAEIAKIEAIOIAPATTIKDQNAQAIATKANERT 1158
 775 NMA---DFTGE---APOGNGENKPSSE-NGKYSTGTVENO-----PTENK 811
 1159 ATAQOTDITAEIAAANAVNDNAVTOANNIEAANSQNDVQAKTGTGEASIDOVTPTVNK 1218
 812 PADSLPEAPNEKPKPENSTDNMG 835
 1219 KATAVTDKANNITAATD---DNGV 1239

RESULT 12
 US-09-815-242-12389
 Sequence 12389, Application US/09815242
 Patent No. US20020061569A1

GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 12389
 LENGTH: 2368
 TYPE: PRT
 ORGANISM: Staphylococcus aureus

US-09-815-242-12389
 Query Match 3.8%; Score 180.5; DB 10; Length 2368;
 Best Local Similarity 19.5%; Pred. No. 0.012;
 Matches 192; Conservative 147; Mismatches 420; Indels 225; Gaps 45;

1 MQITYTDEIQVAKLAGKTYTDEGYIFDTSWIKKDSLEAEAAQAYAKEGLTPPSTD 60
 332 LDLOFIPDNTAVADASRTTNKDGKY-YSFIDNVGLF-----SGSHLYKNRDLAPKATN 386
 61 HQ-----DSGNTAKGAEAIYNRVKAARKVPLDRMPYNLOYTVEVNG---SLIIP--- 108
 387 NKEFTINTEIGNNGNFCASLKADQFKYEVLPGQVTVVNSLTTTPPNGNEDSTVLKMT 446
 109 -SYDHYHNKFEWDEGLYEPKCYSL----DLLATVKYVE-----PRN----- 149
 447 VNYQDAN-KVTFTSQGVTTARGTHTKEVLPDPKSLSKYKVNANIDTPKIDFNKLT 505

Query Match 3.8%; Score 180.5; DB 10; Length 2368;
 Best Local Similarity 19.5%; Pred. No. 0.012;
 Matches 192; Conservative 147; Mismatches 420; Indels 225; Gaps 45;

1 MQITYTDEIQVAKLAGKTYTDEGYIFDTSWIKKDSLEAEAAQAYAKEGLTPPSTD 60
 332 LDLOFIPDNTAVADASRTTNKDGKY-YSFIDNVGLF-----SGSHLYKNRDLAPKATN 386
 61 HQ-----DSGNTAKGAEAIYNRVKAARKVPLDRMPYNLOYTVEVNG---SLIIP--- 108
 387 NKEFTINTEIGNNGNFCASLKADQFKYEVLPGQVTVVNSLTTTPPNGNEDSTVLKMT 446
 109 -SYDHYHNKFEWDEGLYEPKCYSL----DLLATVKYVE-----PRN----- 149
 447 VNYQDAN-KVTFTSQGVTTARGTHTKEVLPDPKSLSKYKVNANIDTPKIDFNKLT 505
 150 ---ASDHYHKNKADQSKDEKDEHSEPHSEDEKENHAGLNPSADNLYKPSSTDE 206
 506 YRTASDIVINN-----AQEVLITADPFSEVAVEMKDALQOQ--VNSQVDNSHYTTASIA 558
 207 E---TEEAEDTDEAIEPGTFSIRQNAETITGLKSSLLLTGKNTISAEDVSLALL 263
 559 EYNKLQOQADNILE--DANHETANRASAADIGLVTKIQALIDQAAIAELD---AKA 614
 264 KESQAPIQGPQIGQPTL-----PNNSLATPSPSLPINTGTSHEKHEEDGYGPOA 313
 615 QKVTAQOQSKKQYQDEVAALVTKINDRNNAIA-----EINKQTT-----A 656
 314 NRITAEDSEGFVMSHSDSNHYFFKDLTEPQIRAKOHLEEVTSNGLDLSLSSHQDYP 373
 657 QGVTEKDNGLAVLDQDVIPTTVKPOAKQDIIQAVTTRKQOIKKSNASLOQDEKDVANDKI 716
 374 G--NAKEMKOLD-----KKTEEK-IAGIMKOYGVKRESIVVNEKKNAILIYPSGDHHH 422
 717 GKIEKAIKIDATTAQVNAEIKTAINDINOTAPATTAKAAALEFDEVQAOIQDAP 776
 423 ADPIDEHKPVGIGHSHSNYELFPEEGVAKKEGNKV-----YTGEELTNVNL 470
 777 LNPDTNEEVAEATERIN-----AAKVSQVKAIEATTAAQDLERVKNEEISKIENI 827
 471 -----LNKSTFN-----NQNTLAN--GOKRVSFSPFPPELEKLGIN--MLVK-- 509
 828 TDSYQTKMDAYNEVQKQATARKTQNAVTSNATNEEVAEADAEEAAQOGLHDIOVYKSK 887
 510 ---LTPDGKVLKVGSKGVGCVGNIANFELDQPYLPQOTPKYTIASKDYPEVSDGTFT 567
 888 QEVADTKSVLDKINAIOQARVPAADTEVENAY-----NTRKQEIQNS-----NASTT 937
 568 VPTSLAY-----KMAQOTIFPPHAGDTYLRVNP-----QFAPVKGTDALVR 609
 938 EEQAATYELDTKKQEAET---NLDAANTNSDVTAKDNGIAAINOVQAAATTKSDAKAE 994
 610 VFDEPHGNAYLENNYKVGEIKLPIPKLNGT---RTAGNKI-PVTFMANAYLDNQSYI 665
 995 IAQ-----KASERKTAIEAMNDSTTEEQAAAKDKYDQAVVVTANADIDNAAAN- 1041

QY 150 ---ASDHVKNKADQSPDEKDEHDEYSEPHSPDEKENHAGLNPSADNLYKSTDTTE 206
Db 506 YRTASDIVINN---AQPEVTLTADPFSVAVEMKNDALQOO--VNSQVDNSHYTTASIA 558
QY 207 E---TEEEAEDTDEAEITPSTIQONAMETITGLKSSLLGTCKDNNTISAEDVDSILLALL 263
Db 559 EYNNKQQADNLINE-DANHVETANRASAADGLVTKLQALIDNQAAIAELD---AKA 614
QY 264 RESQAPAIQGOIGOPTL-----PNNSLATPSPSLPINTGPTSHKHEEDYGDGA 313
Db 615 QEKVTAQSKKVTODEAALVTKINDKNNIA-----EINKQTT-----A 656
QY 314 NRIIADESGFVNSHGDSEHYFFKDLTEQIKAAKHLEEVKTSNGLDLSLSHQDYP 373
Db 657 QGVTEKGNGIAVLQDVITPVTVPQAKODIIOAVTRKQIKKSNASLQDEKDVANDKI 716
QY 374 G---NAKEMKDL-----KIEER-IAGIMKQVGRKRESIVVAKENKAIIPYSGDHH 422
Db 717 GKIEYKAIDDAATNAQVEALKTKAINDINOTAPATAKAAALEEPDEVVQAQIDQAP 776
QY 423 ADPIDHKPVGHSNHYELFKPEGVAKKGNV-----YTGELTNVNL 470
Db 777 LNPDTTNEEVAEAIERIN-----AAKVSQVKAIEATTTAQLDERVKNEEISKIENI 827
QY 471 ---LKNSTFN-----NONFTLAN--GOKRVSPFPPELEKGLIN--MLVK-- 509
Db 828 TDSOTQMDAINEVQKAAATKTONATVSNATNEEVAEADAVERAAKQGLHDQVVKSK 887
QY 510 --LITPDGKVLKVSQVGEVGNIANFELDQYLPFGTKYTIASKDYPEVSYDGTFT 567
Db 888 QEVADTKSKVLDKINAIQVAKVPAADTEVENAY---NTRKQEIQNS-----NASTT 937
QY 568 VPTSLAY-----KMASQTIFFPHAGDYLYLRNP-----QFAPVKGTDALYR 609
Db 938 EERQAAYTELDTKQEAR---NLDAANTNSDVTYAKDNGIAAINOVQAAATTKSKDAKAE 994
QY 610 VFDEFHGNAYLENNYKVEIKLPKLNQCTT---RTAGNKI-PVTFMANAYLDNOSTYI 665
Db 995 IAO-----KASEKTAIEAMNDSTEEQQAARKDVQAVTANADIDNAAN- 1041
QY 666 VEPPILEKENQTPKPSILPOFKRKAQENSKLDEKVEEPRKTS-----EKVEKEKL 715
Db 1042 TDVDNAKTNEATIAITPD--ANVKTAKQAIADKQVQAEATAIDANNAGATTEKAAKQ 1100
QY 716 SETGNSTNSLTLEVPVTVQEKVAKFAESYGNK--LENVLNMGNTIELYLPSGEVIK 774
Db 1101 VQTEKTTADTAIDGAHTNASEVE--AAKNAEIAKIEAIPATTTKDNKQAIATKANERKT 1158
QY 775 NMA---DFTGE---APOGNGENKPSSE-NGKVSTGTVVENQ-----PTENK 811
Db 1159 AIAQTQDITAEETAAANAVNDNAVTOANNIEAANSQNDVQAKTTGEASIDQVTPVTK 1218
QY 812 PADSLEAPNEKPVKPNSTNDNGM 835
Db 1219 KATAVTDAKNNITAATD---DNGV 1239

RESULT 13
US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 3.8%; Score 178; DB 10; Length 2478;
Best Local Similarity 18.9%; Pred. No. 0.018;
Matches 202; Conservative 120; Mismatches 403; Indels 344; Gaps 46;

QY 3 IYTTDEQVAKLAKYITEDYIPTSWIKKDSL---SEAEIRA---AAQAYAKEKGLT 55
Db 1433 VTQIKDQ-AVADIQD---ITADTTIKD---VAKDELATKANEOKALIAOTADATTEEQQA 1486
QY 56 PPSDHO--DSGGTEAKGAEAI-----YNRVKA-----AKVPLDRMP--- 91
Db 1487 NOQVDAQLTQGNQNTENAGSIDDVATKDNAIOADPQASTDVTKNARAELTEMQNKI 1546
QY 92 ---YNLOYTVVEYKNGSLIIPSYDHYHNKFEWFDGLYEAPKGYSGLEDLATVKYIYVPR 148
Db 1547 TELNNNETNEEKGNDIGPVRAAY-----EEGLNNINAATTGQ-VTTAK-----D 1592
QY 149 NASDHVRKNKAQODSKPDEKHEDEVS-----EPHPSDEKENHAGLNPSAD-NLYK 200
Db 1593 TAVQVYQQLHAHPVYKPKAGKELDQAAADKKYQIEQTPNASQOEINDAKQEVDTTELNOAK 1652
QY 201 PSTDTBETEEAEADTDEAEIPCTPSIRONAMETITGLKSSLLGTCKD---NNTISAEDVS 258
Db 1653 TNDVQSSSTNEVYDNAVKECK-----AKINAVKTFSEYKKDALAKIEDAYNAKVNEADS 1706
QY 259 LIALLKESOPAIQGOIGOPTLPPNLSLATPSPSLPIN-----PSTSHEKHE 305
Db 1707 NASTSSEIAEAKQKLAELKQATADQNVNQTAKDDIEVQITHDLDNINDTITPTGKESAT 1766
QY 306 EDGYGF---DANRIIAEDESFGVNSHGDSEHYFFKDLTEQIKAAKHLE-EVKTSHNG 361
Db 1767 TDIYAVADOKKNISADTNA-----TQDEKQAIKQVQDQNVQVTALESING 1812
QY 362 LDSLSSEHQDYPGNAKEMKDLKKIEKTAGIMKQVGRKRESIVVAKENKAIIPYSGDHH 421
Db 1813 VDN-----GDVDDALTQKGAAL---DAIQVDATVTKPKANQAEVKAED-- 1852
QY 422 HADPIDHKPVGHSNHYELFKPEGVAKKGNVYTGELTNVNLKNST----- 475
Db 1853 TKRESIDQSDQL-----TAEKTEALAMIKQITDOAKOG 1885
QY 476 -----FNNQNETLANGQKRVSPFPPELEKKL-----GINMLVKL 510
Db 1886 ITDATTAEVEKAKAOGLEAFDNIQDSTTEKQAIE-----ELETALDQIEAGVYNADA 1940
QY 511 ITPD-----GKVLKVSQVGEVGNIANFELDQYLPFGTKYTIASKDYP 558
Db 1941 TTEKEAFTNALEDILSKATEDISDQTTNAEIAIATVKNASALEQ-----LKAQRINP 1990
QY 559 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDYLYR--VNPQFAPVKGTD----- 606

21.10 2003

1433 VTQKQO-AVADIG--ITADTTIKD--YAKDELAQKANEQKALIAQATADATEEQA 1486
56 PPSDHO--DSGNTAKAEAI--YNRVA--AKVPLDRMP--91
1487 NOQVDAOLTOGNOIENASIDVNTARDNAIOAIDIPQASTOVKTNRABELTEMQNKI 1546
92 ---YNOQVTEVKGSLIPSYDHYHNKPEWDEGLYAPKGYSLLEDLATVKKYVEPR 148
1547 TEILNNETTNEEGNDIGPRAAY--EGLNINAAATTGD-VTTAK--D 1592
149 NASDVRNKNKADQSDPEDKEHDEVS-----EPHPESEKENHAGLNPASAD-NLYK 200
1593 YAVQVQQLHANPVKPKAGKELQAAADKKTOIBOTPNASQOEINDAKQEVDELNOAK 1652
201 PSTDEETEEAEEDTDAEIPGTISRONAMETUTGLKSSLLGTDK--NNTISAEVDS 258
1653 TNVDOSSTNEYDNAVKEGK-----AKINAVTSEYKDKALAKIEDAYNAKVNEADNS 1706
259 LLALLKESQAPITQIGQIPQIOTLPNNSLATPSPSPIN-----PGTSHEKHE 305
1707 NASTSEIAEAKQKLAELKQADQNVNQTAKDDIEVQHNDLNDINDYTIPTGKESAT 1766
306 EDGYGF--DANRIIAEDBSGFMVSHGDSNHFFKDLTEEOIKAAQKHE-EVKTSHG 361
1767 TDLAYADQKKNNISADTNA-----TODEKQOAIKQVDQNVQTALESING 1812
362 LDSLSHEQDYPGNAKEMKDLKIEKTAGIMKQGVKRESIVVKNKERNAIYPSGDHH 421
1813 VDN-----GDVDDALTQGRAAI--DAIQVDATVVKPANOAEVKAED-- 1852
422 HADPIDBHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYGEELTVNVNLLKNT-- 475
1853 TKESIDQDQL-----TAEKTEALAMIKQITDOAKQ 1885
476 -----FNNQNTFLANGKRVSPFPPELEKLI-----GINMLVKL 510
1886 ITDATTAEVEKAKAOGLEAFDNIQIDSTEKQKAE-----ELETALDQAEAGVNVNADA 1940
511 ITPD-----GKVLKSVGKRVGEGVGNIANFELDDQYLPQGTFFYIASKDYP 558
1941 TEEKRAFTNALEDILSKATEDISDQITNABIAVTKNSALEQ-----LKAORINP 1990
559 EVSVDGTFTVPTSLAYKMASOTIFPPHAGDTYLR--VNPQFAVPKGTDA----- 606
1991 EVK-----KNALAEI-----REVYNKQIEIKKNDADASAKAEIART 2036
607 -LVRVDFEFGHNA-YLENNYKVGELK--LPI-----PKLNQGTTRTAGNKIPVTF 652
2027 DLRGYDFRPAKDLKDTQNAEVAELQNTIIPALEAIVPQNDPDANDTNGIDNNDATANS 2086
653 MANAYLDNOSTYIVVEPILEKENOTDKPSILPOFRN-----KA 691
2087 NANATPENTGQ-----PNVSETTANGKADASPTTPNNSDAATGETTATSATDANDKPOA 2141
692 QENSKLDEKVEEPKTSSEKV-EKEKLESTGNSTNSNLEEVPTVDPVQEKVAKFAESYGMK 750
2142 NNNSSVDASTNSPTMDNDVTSKPEVESTNGTIDKPVITEDNATPAESTTNNSTTTATN 2201
751 LENVLNMDGTIELYLPSEVI-----KKNMA--DFTGEAPQNGENKP 792
2202 -ENA-----PTGSTATAPTASTEAASSADSKDNASVNDKQNAEYVNSAESQ 2248
793 SENGKYSTGTVENOPTENK-----PADSLPEAPNEKPKVPKENS----- 830
2249 STNDKVAQPKSENKAKAEKDGSDSTNQSMVESTTETLPISADITEPNVPSNTSKDEESTT 2308
831 ---TDNGMLNPEGVNSD-----PMLDPALEAPAVDPVQEKLEKFTAS 871
2309 NOTDAGQLKSETNVASNEADKSPSKADTEVSNKPSISASSEAKEKMTST 2357

RESULT 15

1 EVK-----KNALAEI-----REVYNKQIEIKKNDADASAKAEIART 2026
607 -LVRVDFEFGHNA-YLENNYKVGELK--LPI-----PKLNQGTTRTAGNKIPVTF 652
2027 DLRGYDFRPAKDLKDTQNAEVAELQNTIIPALEAIVPQNDPDANDTNGIDNNDATANS 2086
653 MANAYLDNOSTYIVVEPILEKENOTDKPSILPOFRN-----KA 691
2087 NANATPENTGQ-----PNVSETTANGKADASPTTPNNSDAATGETTATSATDANDKPOA 2141
692 QENSKLDEKVEEPKTSSEKV-EKEKLESTGNSTNSNLEEVPTVDPVQEKVAKFAESYGMK 750
2142 NNNSSVDASTNSPTMDNDVTSKPEVESTNGTIDKPVITEDNATPAESTTNNSTTTATN 2201
751 LENVLNMDGTIELYLPSEVI-----KKNMA--DFTGEAPQNGENKP 792
2202 -ENA-----PTGSTATAPTASTEAASSADSKDNASVNDKQNAEYVNSAESQ 2248
793 SENGKYSTGTVENOPTENK-----PADSLPEAPNEKPKVPKENS----- 830
2249 STNDKVAQPKSENKAKAEKDGSDSTNQSMVESTTETLPISADITEPNVPSNTSKDEESTT 2308
831 ---TDNGMLNPEGVNSD-----PMLDPALEAPAVDPVQEKLEKFTAS 871
2309 NOTDAGQLKSETNVASNEADKSPSKADTEVSNKPSISASSEAKEKMTST 2357

RESULT 14
US-09-815-242-12967
Sequence 12967, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Cart, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12967
LENGTH: 2478
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 3.8%; Score 178; DB 10; Length 2478;
Best Local Similarity 18.9%; Pred. No. 0.018;
Matches 202; Conservative 120; Mismatches 403; Indels 344; Gaps 46;
QY 3 ITYTDEIQVAKLAKYTTEDGYIFDTSWIKKDSL---SEAPRA---AAQAYAKEKGLT 55

191	QY	LNPSADNLKVPSTDTEE-----TEERAEDTDE--AEI-----PCTPSIRQWAMETLRCGLKSS	241
22	DB	INP-----IIIFAATDSDSLNTDEWEKEETEQPSEVNTGPRIYEAVERSSRDIEKELS	78
242	QY	LLLGTKDNNTISAEVDSLAKALKESQAPIQGPQIGOPTLPNNSLATPS-----	290
79	DB	-----NKVRWNTKAD--LIAMLEKAE--KGPNTNN-----ANSBOTENAAINEEASGAD	124
291	QY	-----PSLPINPGTSHKEHEDEGYGDANRIIAEDESGFVMSHGDSNHYFFKKDL	340
125	DB	RPAIQVERRHPGLPSDSAAEIAKKRR-----ATASSDSELESITYPDKPTKYNNKRV	176
341	QY	TEEQIRAAOKHLEVTSTNGDLS-----LSSHEQD--YPGNAKEKMDLKKTEEKI---	390
177	DB	AKESVADASE--SDLOSSQOSADESSPQLKANOQFPFKVFKKINDAGKVRWDKIDENP	234
391	QY	-----AGIMKOYGVKRESIVVYNNKKNALIIYPSGDHHDAPIDE-----HKPVG	433
235	DB	EVKKAIVDKSAGLIDQLLAKKKS-----EEVNASDFP-----PPPTDEELRLALPETPML	284
434	QY	IGHIS-----HSNVELFKPEEGYAKKEGNKVYTGELTIVNVLLKNSTNNQNFILANGQ	487
285	DB	LGYNAPATSPSPSEFPFPPPP-----TDEEL-----RLALPETPMLGLFNAPATS	328
488	QY	KRYSFSF--PPELEKKLGINMLVKLIITPDGKVLKEVSKVGFGVGNAINFELDQVLPQO	546
329	DB	EPSSEFPFPPPTDELEI-----IRETASSDSSFTRGDLAS-----LRNA	369
547	QY	TFPYTTASKRDYP-----EVSVDGTFVTPTSLAYKMASQTIFFYPFHAGDT-----	590
370	DB	INRHSONFSDPPIPTHEELNGRRGPTSEEFSSSLSGDFTDENSETTEEDIRLADUR	429
591	QY	-----YLRVNPQFA--VPKGTDALVRVDEPHGNAYLENNTKVGEIKLPIPK	635
430	DB	DRGTGKHSRNGFLPLNP--PASSPVLSPKYSKISD-----RALISDIYTKKFPFNPSQP	484
636	QY	LN-----OGTTRTAGNK-----IPVTFMANAYLDONSTYIVYVPILKEKQNOTK	679
485	DB	LNVFNKTTTTTKTPTKPTPVTKAPKLAELPATKPOETVLRENKT-----PFIEKQATNKK	539
680	QY	PSI-----LPQFKKNAQENSKLDEVEEPTSEKVEKELKSTGNSTS-----NSTLEEVP	731
540	DB	QSTINWSPVLYQK-----EATESDKEEMKQOTEENKAYEE--SESANNANGKNRSAGIEE--	591
732	QY	TVDPVQEKVAKFAESYGMKLE	752
592	DB	-----GKLIASAEDEKAKEE	607

GenCore version 5.1.4 ps_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:55:18 ; Search time 36 Seconds
(without alignments)
3332.506 Million cell updates/sec

Title: US-09-884-465A-332
Perfect score: 4728
Sequence: 1 MQITYTDEIQVAKLAGKYP.....IELRLPSGEVKNLSDFIA 906

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 619202 seqs, 132417472 residues

Total number of hits satisfying chosen parameters: 619202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pap.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pap.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pap.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pap.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pap.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pap.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1302.5	27.5	838	5	US-09-468-656A-4
2	937.5	19.8	819	5	US-09-468-656A-10
3	818.5	17.3	763	5	US-09-765-272A-66
4	637	13.5	919	5	US-09-468-656A-8
5	636	13.5	796	5	US-09-765-272A-56
6	470	9.9	484	5	US-09-468-656A-6
7	431	9.1	447	5	US-09-765-272A-182
8	331.5	7.0	485	5	US-09-769-736-72
9	223.5	4.7	822	5	US-09-769-736-18
10	223.5	4.7	6641	6	US-10-282-122A-70580
11	223.5	4.7	10382	6	US-10-092-411A-3159
12	201	4.3	1790	6	US-10-369-493-1586
13	199.5	4.2	1098	5	US-09-797-385-8
14	197	4.2	1478	6	US-10-369-493-22154
15	195.5	4.1	1128	5	US-09-797-385-2
16	194.5	4.1	1164	5	US-09-797-385-6
17	193	4.1	1245	6	US-10-172-502-19
18	192	4.1	571	5	US-09-765-272A-4
19	192	4.1	1164	5	US-09-797-385-10
20	192	4.1	2397	6	US-10-282-122A-71232
21	191.5	4.1	3692	6	US-10-282-122A-71235
22	191.5	4.1	3696	6	US-10-092-411A-5080
23	190	4.0	4688	6	US-10-282-122A-76865
24	189.5	4.0	5005	6	US-10-282-122A-76871
25	185.5	3.9	1104	5	US-09-797-385-4
26	185.5	3.9	2109	6	US-10-369-493-6346

27	185	3.9	1633	6	US-10-282-122A-70437	Sequence 70437, A
28	184.5	3.9	1963	6	US-10-282-122A-73978	Sequence 73978, A
29	183.5	3.9	972	6	US-10-282-122A-53353	Sequence 53353, A
30	181.5	3.8	711	6	US-10-341-201-3	Sequence 3, Appl
31	181.5	3.8	1139	6	US-10-282-122A-63564	Sequence 63564, A
32	181	3.8	2481	6	US-10-282-122A-43762	Sequence 43762, A
33	181	3.8	2495	5	US-09-950-084-6143	Sequence 6143, Ap
34	180.5	3.8	1183	5	US-09-751-708A-45	Sequence 45, Appl
35	180.5	3.8	2045	6	US-10-282-122A-74463	Sequence 74463, A
36	178.5	3.8	1946	6	US-10-282-122A-62947	Sequence 62947, A
37	178	3.8	1920	6	US-10-282-122A-71413	Sequence 71413, A
38	177.5	3.8	1665	6	US-10-282-122A-71690	Sequence 71690, A
39	176	3.7	885	6	US-10-282-122A-46439	Sequence 46439, A
40	175	3.7	1183	6	US-10-092-411A-3530	Sequence 3530, Ap
41	174	3.7	1270	1	PCT-US02-38445-39	Sequence 13, Appl
42	174	3.7	3957	6	US-10-228-207-13	Sequence 39, Appl
43	173	3.7	2840	7	US-60-443-566-4237	Sequence 4237, Ap
44	173	3.7	3957	7	US-60-443-566-4236	Sequence 4236, Ap
45	171.5	3.6	916	5	US-09-724-676-92902	Sequence 92902, A

ALIGNMENTS

RESULT 1

US-09-468-656A-4
; Sequence 4, Application US/09468656A
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-4

Query Match 27.5%; Score 1302.5; DB 5; Length 838;
Best Local Similarity 81.6%; Pred. No. 7.9e-87;
Matches 262; Conservative 2; Mismatches 6; Indels 51; Gaps 4;

QY	2	QITYTDEIQVAKLAGKTYTDEGYIFD	-----TSWIKKDSLEAEAR 42
DB	518	QITYTDEIQVAKLAGKTYTDEGYIFDPRDITSDGDAVYTPHMTSHWIKKDSLEAEAR 577	-----TSWIKKDSLEAEAR 42
QY	43	AAQAYAKEGLTPSTDHQDSGNTGAEAIYRVRAAKVPLDRMPYNLOYTVYVKN 102	-----
DB	578	AAQAYAKEGLTPSTDHQDSGNTGAEAIYRVRAAKVPLDRMPYNLOYTVYVKN 637	-----
QY	103	GSLLIPSYDHYNIKFEWDEGLYEPKGYSLDLELATVYKYYE-----PR-----NAS 151	-----PR-----NAS 151
DB	638	GSLLIPSYDHYNIKFEWDEGLYEPKGYSLDLELATVYKYYEHPNPHRPHSDNGFGNAS 697	-----PR-----NAS 151
QY	152	DHYVRKNADQSKPDEKDEHDEVPSETHPESEKENHAGLNPSADNLKYPSTDTETETEE 211	-----PR-----NAS 151
DB	698	DHYVRKNADQSKPDEKDEHDEVPSETHPESEKENHAGLNPSADNLKYPSTDTETETEE 757	-----PR-----NAS 151
QY	212	AEDTTDEAEIPQVSVINAKIADAEALKEKVTDPISQIRONAMETLTGLKSSLLGKTKNN 250	-----PR-----NAS 151
DB	758	AEDTTDEAEIPQVSVINAKIADAEALKEKVTDPISQIRONAMETLTGLKSSLLGKTKNN 817	-----PR-----NAS 151
QY	251	TISAEVDSLLALKESQAPPI 271	-----PR-----NAS 151
DB	818	TISAEVDSLLALKESQAPPI 838	-----PR-----NAS 151


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272A
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Lin J. Hymel
REGISTRATION NUMBER: 45,414
REFERENCE/DOCKET NUMBER: PB340P2C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 610-5790
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272A-66
Query Match 17.3%; Score 818.5; DB 5; Length 763;
Best Local Similarity 66.1%; Pred. No. 2.1e-51;
Matches 168; Conservative 16; Mismatches 29; Indels 41; Gaps

QY 2 QITYTDEIQVAKLAGKTYTDEGTYFD-----TSWIKKDSLSEAE 42
DB 498 QITYTDEIQVAKLAGKTYTDEGTYFDPRDTSDEGDYATPHTSHWIKKDSLSEAE 557
QY 43 AQAQYAKKGLTPSTDQDSGNTKAGAEAIYNRVKAAKKVPLOMPYNLOYTYVEKN 102
DB 558 AQAQYAKKGLTPSTDQDSGNTKAGAEAIYNRVKAAKKVPLOMPYNLOYTYVEKN 617
QY 103 GSLIIPSYDHYHNKIFWFDGLIYAPKGYSLDLLATVKYTYE-----PR-----NAS 151
DB 618 GSLIIPSYDHYHNKIFWFDGLIYAPKGYSLDLLATVKYTYEHPNRPSPHSDNGFGNAS 677
QY 152 DHYRKNK-----ADQSKPKDEKDEHVESEPTHPSEDSKENHAGLNPSAONLYKPSDTDE 206
DB 678 DHVRNKNQADNTQTEKSEKPKTEKPEETPREKPKQSEKPEP-----KPTPEE 731
QY 207 ETEEEAEDTTDEAE 220
DB 732 ESPESEEPQVETE 745

RESULT 4
US-09-468-656A-8
; Sequence 8, Application US/09468656A
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRP
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-8
Query Match 13.5%; Score 637; DB 5; Length 819;
Best Local Similarity 45.0%; Pred. No. 4.7e-38;

```

Matches 140; Conservative 41; Mismatches 76; Indels 54; Gaps 8;

QY 2 QITTTDEIOVAKLAGKTYTDEGYIFD-----TSWIKKDSLSAEAR 42
 Db 527 QIETTEDVRIAQLADKRTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLSDEK 586
 QY 43 AAOAYAKEGLTPSPDQSDGNTAEKAGAEAIYNRVKAARKVPLDRMPYNLOTYTVEVKN 102
 Db 587 VAAQAYTKEGILPPSPDADVKANPTGDSAAAIYNRVKEGRIPVLRPLVYVHEVTEVKN 646
 QY 103 GSLIIPSDHYHNIKFEWFDGLYAPKGYSLDGLLATVKKYVE-----PR-----NAS 151
 Db 647 GNLIIHPKHDIHNIKFAWFDDBHYKAPNGYTLEDLFATIKYIVHPDERPHSNDGNGNAS 706
 QY 152 DHVRKNKADQSKPDDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPDSTDEETEE 211
 Db 707 EHV-LGKKHSDPNKFKADE--EPVEETPAEPE-----VPQVETEKVEAQ 750
 QY 212 AEDTTDEAEI-----PGTSPSRQNAOMETLTGLKSSLLGTCKDNNTISAEVDSLALLKESQ 267
 Db 751 LK-----EAEVLLARVTDSSLKANATETLAGLRNLTQIMDNNSIMAEKLLALLKGSN 806
 QY 268 PAPIOGPOIGQ 278
 Db 807 PSSVSKEKINK 817

RESULT 5
 US-09-765-272A-56
 ; Sequence 56, Application US/09765272A
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 454
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: Dell Latitude C610
 ; OPERATING SYSTEM: Windows 2000
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272A
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: OCT-30-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lin J. Hymel
 ; REGISTRATION NUMBER: 45,414
 ; REFERENCE/DOCKET NUMBER: PB340P2C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 610-5790
 ; TELEFAX: (301) 309-8439
 ; INFORMATION FOR SEQ ID NO: 56:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 796 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
 US-09-765-272A-56
 Query Match 13.5%; Score 636; DB 5; Length 796;
 Best Local Similarity 45.3%; Pred. No. 5.3e-38;

Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps 8;

QY 2 QITTTDEIOVAKLAGKTYTDEGYIFD-----TSWIKKDSLSAEAR 42
 Db 507 QIETTEDVRIAQLADKRTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLSDEK 566
 QY 43 AAOAYAKEGLTPSPDQSDGNTAEKAGAEAIYNRVKAARKVPLDRMPYNLOTYTVEVKN 102
 Db 567 VAAQAYTKEGILPPSPDADVKANPTGDSAAAIYNRVKEGRIPVLRPLVYVHEVTEVKN 626
 QY 103 GSLIIPSDHYHNIKFEWFDGLYAPKGYSLDGLLATVKKYVE-----PR-----NAS 151
 Db 627 GNLIIHPKHDIHNIKFAWFDDBHYKAPNGYTLEDLFATIKYIVHPDERPHSNDGNGNAS 686
 QY 152 DHVRKNKADQSKPDDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPDSTDEETEE 211
 Db 687 EHV-LGKKHSDPNKFKADE--EPVEETPAEPE-----VPQVETEKVEAQ 730
 QY 212 AEDTTDEAEI-----PGTSPSRQNAOMETLTGLKSSLLGTCKDNNTISAEVDSLALLKESQ 267
 Db 731 LK-----EAEVLLARVTDSSLKANATETLAGLRNLTQIMDNNSIMAEKLLALLKGSN 786
 QY 268 PAPIOGPOI 276
 Db 787 PSSVSKEKI 795

RESULT 6
 US-09-468-656A-6
 ; Sequence 6, Application US/09468656A
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 484
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-468-656A-6
 Query Match 9.9%; Score 470; DB 5; Length 484;
 Best Local Similarity 33.6%; Pred. No. 3.6e-26;
 Matches 131; Conservative 40; Mismatches 91; Indels 128; Gaps 12;

QY 7 DDEIOVAKLAGKTYTDEGYIF-----DTS-----WIKKDSLSAEARAAQA 47
 Db 168 NSNVAVSQGRYITNDGYTFNPADIETGNAYVPHGGHYHYPKSDLSASELAAGA 227
 QY 48 YAKEGLTP-----PSTDHQDSGNTAEKGA-----EAIYNRVKAARKVPL 87
 Db 228 HLAGKNQPSQISYSSTASDNTQSVANGSTKPAKNSLQSLKELYDSPSAQRYSES 287
 QY 88 DRMPYNLOTYTVEKNGSLIPSDYHNIKFEWFDGLYAPKGYSLDGLLATVKKYVEP 147
 Db 288 DGLVFPDAKIISRTPNQVAIPHGDHYHFPYS-----KLSALEEKIARM---VPI 334
 QY 148 RNASDHVRKNKADQSKPDDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPDSTDEE 207
 Db 335 SCTGSTVSTN-----AKPNEV-----SSLGSLSN---PSSLTIS 367
 QY 208 TEEAEADTTDEAEIPCTPSIRQNAOMETLTGLKSSLLGTCKDNNTISAEVDSLALLKESQ 267
 Db 368 KE-----LSSASDGYIFNPKD-----IVEETA 389

41.11 2003

QY 208 TEEAEADTTDAEIPGTPSIRONAMETILGKSSLLGLTKDNNTISAEDVSLALLKESQ 267
 Db 344 KE-----LSSASDGYIENPKD-----IVEETA 365
 QY 268 PAPI-----QGPOIGQPTLPNNSLATPPSPSLINPOTSHKHEEDGYGFDANRI 316
 Db 366 TAYIVRHGDHFHYIPKSNQIGQPTLPNNSLATPPSPSLINPOTSHKHEEDGYGFDANRI 425
 QY 317 IADESGFVMSHGSDSNHYFFKK 338
 Db 426 IADESGFVMSHGSDSNHYFFKK 447

RESULT 8
 US-09-769-736-72
 ; Sequence 72, Application US/09769736
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Le Page, Richard WF
 ; APPLICANT: Wells, Jeremy M
 ; APPLICANT: Hanniffy, Sean B
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21089WC
 ; CURRENT APPLICATION NUMBER: US/09/769,736
 ; CURRENT FILING DATE: 2003-02-14
 ; PRIOR APPLICATION NUMBER: GB 9816335.5
 ; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/125163
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 212
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 72
 ; LENGTH: 485
 ; TYPE: PRT
 ; ORGANISM: Streptococcus agalactiae
 US-09-769-736-72

Query Match 7.0%; Score 331.5; DB 5; Length 485;
 Best Local Similarity 23.8%; Pred. No. 5.2e-16;
 Matches 95; Conservative 48; Mismatches 89; Indels 167; Gaps 9;
 QY 2 QTYTDDDEIQVAKLAGKYTTEDGYIFDTS-----WIKKDSLSAEAR 42
 Db 94 QIVYSAQIEEAKKAGYTTSDGYIFDADKKDKDTGTGVIPHMTHEHWVPKDLSESEL 153
 QY 43 AAAQAVY-----49
 Db 154 KAAQEFSLKSEANQDKPKTKTAQIEVEAIEPKAIVKPEDLLFGIAQATYKNGTFVIP 213
 QY 50 -----KEKGL-----54
 Db 214 HKDHYHYVELKWFDEBKDLADSDKTSYLEDYLATAKYMMHPRKPRKVEGKDAEYK 273
 QY 55 -----TPSTDHODSGNTEAKGAZA--IYNNRKAARKVPLDRMPYINQYTVKN 102
 Db 274 EKDSNKADKPSAPTNDKSTNSGDKNLKSAEAFKQAKPEKIVPLDKIAAHMAYAVGED 333
 QY 103 GSLIIPSYDHYHNIKFEWFE-GLYEAPKGYSLDGLATVYVEPRNA-----S 151
 Db 334 DQIVPHDHYHNVPMWFDKGLKAKPEGYTLQQLFSTIKYIMEHPNLPKKGWGHDS 393
 QY 152 DHYRKNKADQSK---PDEKHEDEVSEPHPESEKENHAGLNPSADNLYKPSDTEET 208
 Db 394 DHNKGSKMKNKANYAPDEE-----PEDSGKVTH---NYGFYDVNKGSDDEEPE 439
 QY 209 EEEAEADTTDAEIPGTPSIRONAMETILGKSSLLGLTK 247
 Db 440 KOESELDELTELMQNAKAKYGNDRQSFKEQLIQLSNK 478
 RESULT 9
 US-09-769-736-18
 ; Sequence 18, Application US/09769736

QY 317 IADESGFVMSHGSDSNHYFFKKOLTEQIK 346
 Db 450 IADESGFVMSHGSDSNHYFFKKOLTEQIK 479

RESULT 7
 US-09-765-272A-182
 ; Sequence 182, Application US/09765272A
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
 ; NUMBER OF SEQUENCES: 454
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: Dell Latitude C610
 ; OPERATING SYSTEM: Windows 2000
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272A
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: OCT-30-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lid J. Hymel
 ; REGISTRATION NUMBER: 45,414
 ; REFERENCE/DOCKET NUMBER: PB340P2C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 610-5790
 ; TELEFAX: (301) 309-8439
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
 US-09-765-272A-182

Query Match 9.1%; Score 431; DB 5; Length 447;
 Best Local Similarity 32.2%; Pred. No. 2.3e-23;
 Matches 123; Conservative 40; Mismatches 91; Indels 128; Gaps 12;
 QY 7 DDEIQVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSAEARAAQA 47
 Db 144 NSNVAARSQRYTNDGYVFNPAIDEDTGNAYIVPHGHHYIPKSDLSASELAAGA 203
 QY 48 YAKEGLTP-----PSTDHODSGNTEAKGA-----EAYNNRKAARKVPL 87
 Db 204 HLAGKNMOPSQLSYSTASDNTOSVAKGSKPANKSENQSLKELYDSPSQRYSES 263
 QY 88 DRMPYNLYTVKNGSLIIPSYDHYHNIKFEWFEGLYEAPKGYSLDGLATVYVEPR 147
 Db 264 DGLVDPDKAIISRTPNGVAIPGHGHHYIPY-----KLSALEEKIARM---VPI 310
 QY 148 RNASDHYRKNKADQSKPDEKHEDEVSEPHPESEKENHAGLNPSADNLYKPSDTEET 207
 Db 311 SGTGSTVSTN-----AKPNEV-----SSLSGLSSN---PSLSTTS 343

```

; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard Wf
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PNC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-18

Query Match 4.7%; Score 223.5; DB 5; Length 822;
Best Local Similarity 19.0%; Pred. No. 9.8e-08;
Matches 186; Conservative 143; Mismatches 319; Indels 301; Gaps 52;

QY 65 GNTKAGAEATYVRKAAKVPDRMPYNLQYTVKNGSLIIPSDYDHYH-----114
DB 47 GKVKAPKTKNTWDOISAEGISAB-----QIVVKITDGYTVSHGDHYHFNKGVKPYDA 100
QY 115 -----NIKPFWDEGLYAPKGYSLIEDLATVYIVY--EPRNASDHVRKMK--A 159
DB 101 IISEELLMTDPNTHPKQSDV--INEILDGVIVK---VNGNYVTVKPKSKRNKIRTKQIA 156
QY 160 DQSKPDEKDEHDEVEPHPSDEKENAGLNPSADNLKYSTDTETEEREAEDTTDEA 219
DB 157 EQVAKTEKAKERGLAQVAHL-----SKEEVAVN-----EAKRGYRTDGG 199
QY 220 EIPGTPSIRONAMETITGLKSSLL-----LGTKNNTISAEDVSLALLKESQ--- 267
DB 200 YIFSPTDI-----IDLDGDAYLVPHGNHYHYPKKDLS--PSELAAQAQYWSQKRG 250
QY 268 -----PAPIQG-----PQIGQPTLPNNSLATPSLPIPNPGTSHSEKH----- 304
DB 251 ARPSDYRPTPAPGRRKAPIDVTPNPGQHQPDNGGYPHAPRPND--ASQNKHQRDEPK 308
QY 305 -----EDGKGFDANRIIAEDSGFVMSHSDSNHYFFKKDLTEE 343
DB 309 GKTFKELLHLRLDLKYRIVEEDGLIPEPTQVIKSNAGYVYPHGDHYHILPRSLSPL 368
QY 344 QIKAAQKHLEEVKTSNGLSLSHEQDYPGNNAKEMKDLDKKTEEKIAG--IMKQYGV--- 399
DB 369 EMELADRYLAG-QTDON--DSGSDHS-----KPSDKVETHFTFLHRIKAYKGLD 415
QY 400 -----KRESIVNKE-----KNALYPSGDHHDHAPIDEKKPVGIGHSHSNTELEKPEE 448
DB 416 GRPYDTSAYVFSKESHSYKSGVTAKEGDHPHY-----IGFGEI--EYELDEVAN 466
QY 449 GVAKEGKNYVTGEBELTNVNLKNSTFNQNTLANGOKKVSFSPPELEKLGIMLV 508
DB 467 WV-KAKGQ-----ADELVAUD-----QEGKKEKPLFD-----TKKVS----- 498
QY 509 KLITPDGKV--LEKVSQKVFGEVGNIANFELDQPLPGQTFRKTYTASQDPEVSYDGT 566
DB 499 RYVTKGQKGVIMPDKGKY-----PYARYQLDLTQI--AFAEQELMLKDKHYRYD--- 548
QY 567 TVPSTSLAYKASQTIPTFFIAGDTYLRVNPQPAVPGKTDALRVFDEFHGNAY--LENNY 624
DB 549 IVDTGIEPRLAVDSLSLPMAGNATYDTGSSFVIPH-----IDHIVVYPYSWLTNRN- 599
QY 625 KVGEEKLPIPLKNGTTRTAGNKIPVTFMANAVLDNGSTVIVEVP-----ILEKENQTDK 679
DB 600 QIATIK-----YVMQHPVRRPDVWSKPGHEES 626

; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70580
; LENGTH: 6641
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: MISC-FEATURE
; LOCATION: (6636)..(6636)
; OTHER INFORMATION: X-any amino acid
US-10-282-122A-70580

Query Match 4.7%; Score 223.5; DB 6; Length 6641;
Best Local Similarity 19.0%; Pred. No. 2.6e-06;
Matches 201; Conservative 169; Mismatches 409; Indels 279; Gaps 47;
```

us-09-884-465a-332.rapn

.1:11 2003

```
RESULT 11
US-10-092-411A-3159, Application US/10092411A
Sequence 3159, General Information:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-101
CURRENT APPLICATION NUMBER: US/10/092,411A
CURRENT FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: US 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5676
SEQ ID NO 3159
LENGTH: 10182
TYPE: PAT
ORGANISM: Staphylococcus epidermidis
US-10-092-411A-3159

Query Match 4.7%, Score 223.5, DB 6; Length 10182;
Best Local Similarity 19.0%, Pred. No. 5.1e-06;
Matches 201; Conservative 169; Mismatches 409; Indels 279; Gaps 47;

QY 37 LSEARAAQAYAKEKGLTPPSTHDSGNTAKGAEATYNNRVKAAKVPLODRMPYLNQ- 95
1704 LTRAKEDAVASINNLGTLNEQPKENAVGAQTRDOVANKLRDAE--ALDQSMOTLRD 1761
QY 96 -----YTVVEKNGSLIPSYDHYHNIKFWEFDEGL----- 125
1762 LVNNQNAIHSTSYFNEDSTQKNTYDAIDNGSTYITG-QHNPELNKSTIDQISRINTA 1820
QY 126 -----YEAPKGYSLDGLATVKKYVPERNASDVRKNKADQDSKPDE----- 167
1821 KNDLHGVEKLQDKGTANOE-IGOLGYLNDPQKSGEESLYNGSNTSRSEVHEHLNEAKSLN 1879
QY 168 ---DKEHDEVESEPHPE-----SDEKENHAGLNPS---ADNLY-----KPSDTDEETEEE 211
1880 NAMQLRDKVAEKTNVKSSDYINDSTEHQGYDQALAEANIINEIGNPTLNKSEIQK 1939
QY 212 AEDTDEAEIPGTPSIRONAMETLTGLKSSLLGTGKNTTISAEDVSLALAKESQAPAI 271
1940 LQQLTD-----AQNA-----LOGSHLLEAKNNAITG-INKLALDAQORQAI 1982
QY 272 QGPGIQGQPTLP-----NNSLATPSPSLPINPGTSHEKHEEDGY-----G 310
1983 ENVO-AQOTIPAVNOQLTLDRINTAMQALRKVQGNVHQSNTYFNEDEQPKHNYDNS 2041
QY 311 FDNRIIAEDESFGVMHSGDSNHYFFKDLTEQIKAAOK-HLEEVKTSH--NGLDLSL- 366
2042 VQAGQTIIDKLQDPIMNKNEIEQAINQINTOTALSGENKLTQDSTNRHQIEGLSSLT 2101
QY 367 ---SHEQDYPGNAKEMKDLKKI-----EKKIA-----GIMKQYGVKRESIVVYKKNKAI 413
2102 AQINAEKDLVQAKTRTDVAQKAAAEINSANSLRDGQONKEDIKRSAYINADPTKV 2161
QY 414 IYPSGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKVTYTGELTNVYLLKN 473
2162 TAYDQALQNAENIINATP-----NVELNK-----ATIEQALSrvQQAQDLDGVQOLAN 2210
QY 474 S-----TFNNQNTFLANGQKRVSFSPPELEKLGINMLVKLTIPDGKLVKESG----- 523
2211 AKQATQTVNGLN-SLNDGQKR-----ELNLLINSANTRTKVOEELNKATEL 2256
QY 524 ---KVFGEVGNINANFELDQPLPGQTFKTIASKDYPEV-SYDGTFTVPTSLAYKNAS 578
2257 NEAMEALNSVQNVDOVKQSSNYV-----NEDQPEQHNYDN-----AVNEAQ 2298
QY 579 QTFYFFPHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGHNAVLENNYKVGEIKL----- 631
2299 ATI-----NNAQVLDKLAIERLTQVTNTTKDALHGAQKLTQDQAAETGIRGLTSLN 2352
QY 632 -----PIPLKNGQTRTAGNKI--PVTMANAYL-----DNQSTYIVE----- 667
2353 EPQKNAEVAKYAATTRDEVRNIRQEAATLDTAMLGLRKSIGKDKNTKSKYINEDHDQ 2412
QY 668 -----VPILEKQNTDPSILPQFKRNKAQENSKL--DERVEEPTKSEK---VE 711
2413 QQAYDNVAVNAQVIDEOTATLSSDTINQLANAVTQAKSNLHGDTKLQHKDSAKOTIAQ 2472
QY 712 KEKLSGTGNTSNTLEVTVPVQKFAEYSGMKLENVLFNMDGTIELX---LPS 768
2473 LQNLNSAQKEMEDSLIDNESTRITQVQHDLE-----AQALDGLMGALKESIKDITNIVSN 2527
QY 769 GEVI-----KKNADFTGEAPQG---NGENKPSN-KGVSTG--TVENOPTENKPADSLPE 818
2528 GNYINAEPSKQAYDAVQNAQNIINGNTQNTKNGNVITATQVKNYTKDALDGDHRLLE 2587
QY 819 APN---EKVPKVENSTDNGLMPEGVGSDPMLDPALEAPAVDPVQEKLEKFTASYIGL 876
2588 ARNNAHOTIRNLNLSNNAQKDAEKL-----VNSASTLEQVQOOLN--TAQQ---L 2633
QY 877 DSVIFNMDGTIELR-----LPSGEVIRKNLSDFI 905
2634 DNANGELQSTIAKQOVKADSKYLNEDPQIKQNYDDAV 2671
```

Db 5867 ATI-----NNNAOVLKLAIERLTQTNTTKDALHGAQKLTQOQAAETGIRGLTSLN 5920
 QY 632 -----PIPLNGOTRTAGNKI--PVTMANAYL-----DNQSTYIYE-----667
 Db 5921 EPORNAEAVKATTAATRDVNRNRQBATLDTAMUGLRKSIKDKNTKSSKIYINEDHQ 5980
 QY 668 -----VPILEKQNTDPSLPKRNKAQENSKL--DEKVEPKYSEK---VE 711
 Db 5981 QOAYDNVNAOQVIDETQATLSSDTINOLANAVTOAKSNLHGDTKLOHDKDSAKOTIAQ 6040
 QY 712 KEKSETGNTSNTLEEVTPDQVQKPAESYGMKLENLFWNDGTIELY---LPS 768
 Db 6041 LONLSAQKMHESLIDNSTRTOVQHDLTE-----AQLDGLMGALKESIKDTYTNVSN 6095
 QY 769 GEV-----KKNMADTGEAPOG--NGENKPSN--GKYSTG--TVENQPTENKPADSLPE 818
 Db 6096 GNYINAEPKQAYDAVQNAQNIINGTNPNTINKGNVTATQTVKNTKDALDGDHRL 6155
 QY 819 APN--EKPVKPSNDGMLNPGVSGDPMPLDPALEAPADVPVQKLEKFTASYGLGL 876
 Db 6156 AKNNANQTIIRNLNNAQDAEKNL-----VNSASTLEQVOOQLQ--TAQO---L 6201
 QY 877 DSVTFNDGTIELR-----LPSGEVKKNLSDPI 905
 Db 6302 DNAMGELRQSTAKKQVKAQSKYLNEDPQIKQNTDDAV 6239

RESULT 12
 US-10-369-493-1586
 ; Sequence 1586, Application US/10369493
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 1586
 ; LENGTH: 1790
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-369-493-1586

Query Match 4.38; Score 201; DB 6; Length 1790;
 Best Local Similarity 19.36; Pred. No. 1.5e-05;
 Matches 183; Conservative 132; Mismatches 328; Indels 304; Gaps 40;
 QY 67 TEAKGAEIYNRVKAQVPLDRMPYNLQYTVGVKNGSLIIPSYD-----HYHNKFE 119
 Db 650 TTKLGNVYASIKQFKK-----DSYFSDVNMEDSILAPDELDTGLPKYVFSTYFIQ 702
 QY 120 WFDGLYEAAPKGYSLDILLATVYVYVPRNASHVYKKNKADQSPDDEKHEDEV-----174
 Db 703 LFENIYRI-----RTALSH-----DPDEEPINKISFEEVERKLOR 737
 QY 175 -----SETHPESDEK-----ENHAGLNPDSADNYKPKSTDEE 207
 Db 738 QCTKLAGEITSQTETSTHENTKLTALTHEHDELKYOILNASSHSL-KENFSILE 796
 QY 208 TE-EEAEDTTDEAIEPTGSIRQNAEMETLGLKSLLLGTDKNNTTISAFVDSLALLKES 266
 Db 797 TELKNYRDSILDE-----MTQLRDVLETKDKENQTALEYKSTI-----HK 836
 QY 267 QPAPIQGPQIGPGLPNNSLATPSPSLPNTPGTSHKHEEDG-----YGFANDRIIA 318

Db 837 QEDSKITLEKLETI-----LSQKKKAEDGINKMGKDLFALSREMQAV 879
 QY 319 EDESGFVMSHDS---NHFFKKDLTEB-----QITAAQKHLEEVKTSNGL-----DSL 365
 Db 880 EENCKNLQKEDKSNVNHQKETSLEKEDIAAKITEIKAINLENLEEMKIQCNLSKEKEHI 939
 QY 366 SSHEDQYPGNAKEMKDLKKIEEIKIAGIMQY---GVKRESIV--VNKEKN--AIYPSGD 419
 Db 940 SKELVEYKRSQSHDNLVAKLTEKLSLANNYKQMAENESLIRAVEESKNESQIOLSNL 999
 QY 420 HHADPIDEHK---PVGIGHSNVELFK-----PEGVAKKEGNKYVTEELTNV 467
 Db 1000 ONKIDMSQERENFOIERGSEIKETKTIISLEQTKKEEIIKSDSSK-----DEYESQ 1055
 QY 468 VMLLN-----STFNNQNTLANGOKRYSFSPPELE---KKLGINMLVKLITPGKVLEK 520
 Db 1056 ISLLKEKLETATANDENVNKISELTKTRELEAEALAYKNLKNLETKLETSE-KALKE 1114
 QY 521 V-----SGKYFGEVGNIANFELDQPLPGQTPKY---TIASKOYP 558
 Db 1115 VKEZEHLKEEKIQLEKEATEKQQLNSLRANLESLEKEHEDELAALAKYEEQIANK---1171
 QY 559 EYSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQF-AVPKGTDALVRVDFEFGN 617
 Db 1172 ERQYN-----EISQLNDEITSTOOENESIKKNDLEGE 1206
 QY 618 AVLENNYKVGELKPIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIYVEVPILEKQNT 677
 Db 1207 V-----KAMKSTSEQSN-----LKKSEIDALNLOKE---LAKKNET 1241
 QY 678 DKPSILPOPK-----RNKAQENSKLDEKVE---EPKTEKVEKEKLESE---717
 Db 1242 NEASLLESIKSVESVTVKIKELQDECNFEKEVSELEDKLASEDKNSKYLEKSEKI 1301
 QY 718 -----TGNSTSNSTLEEVTPDVPQEK-----VAKFAESYGMKLENVLFNM 758
 Db 1302 KEELDAKTELKIQLEKINTLAKAKESSELSRLKTSSEKKNABEQLEKLN-----1356
 QY 759 DGTIELYPSGEVYKKNMADTGEAPQNGENKPSNGKSTGTVENQPTENKPADSLPE 818
 Db 1357 -----ELQIKQA-----FEKERLLNEGS-STITOYSEKINTLEDELIR 1396
 QY 819 APNEKVPKPSNDGMLNPGVSGDPMPLDPALEAPADVPVQEK 865
 Db 1397 LONENELKAKEIDNTRSELEKVSLSNDELLE-----EKONTIKSLQDEI 1440

RESULT 13
 US-09-797-385-8
 ; Sequence 8, Application US/09797385
 ; GENERAL INFORMATION:
 ; APPLICANT: Tal, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: Non-IgA Fc Binding Forms of the Group B
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC Compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/797,385
 ; FILING DATE: 01-Mar-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/923,992
FILING DATE: 05-SEP-1997
APPLICATION NUMBER: US 60/024,707
FILING DATE: (VII)

ATTORNEY/AGENT INFORMATION:

NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1098 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-797-385-8

Query Match 4.28; Score 199.5; DB 5; Length 1098;

Best Local Similarity 17.69; Pred. No. 8.8e-06;

Matches 157; Conservative 125; Mismatches 289; Indels 321; Gaps 34;

QY 166 DEKHEVSEPTHP---ESDEKENHAGLNPSADNLYKPSDTTETETEEAEDTDEAIP 222

Db 12 DSVKTEVAAPYPSMACTDQGNSSSELETRMEIPTTDIKKAVEPVETAGETSAT 71

QY 223 GTPSIRQAMETLTGLKSL---LIGTRDNFTISAIV---DSLLAL----- 262

Db 72 HTGKREKQLQOMKNNKNDVDNTILSHQENFEKTKIDTNDSDALLEENGFNETNRL 131

QY 263 -LKESQAPIQGPIQPLNNSLATPSPINPGTSHKHEEEDGYGFANDRIADE 321

Db 132 HIKOHEVEEKAKAOKVILKQSD---TKVDSLNDIKELNHQKSQAG-----ITNEDK 182

QY 322 SGVMSHGDSNHYFFKKDTE-----EOLK----- 346

Db 183 DSNLKTIEDIRKQAQPDKEDAEVREELKLFSTKAGLDQEOEHVKETSSEWT 242

QY 347 -----AAOKHLEVTSHGLDLSLSHQ-----DYPGNKEMKDLKKIE 387

Db 243 QKDEHYANSLONLAQSLSE-----LDKATNEQATQVKQFLENKAKLEIOLPIK 295

QY 388 EKIAIMKQYGVKRESI-----VYNNKENALIYPSGHHHADPIDEKPVG 433

Db 296 ETNVLKAMSELESQVEKELKHNSANEDLVAKSREIYREYEGKLNQSKNLPKQLE 355

QY 434 IGHSHNVELEKPEGVAKKEGKRVYTGELTNVNLKSTNNONFTLANGKRVSF 493

Db 356 -EAAHSL-----KQVVEOPRKKFKTSEQVTPKKRLKRLAANE-----NQOKIELT 402

QY 494 FPELEKLGINMLVKLITPDGKLVKRVGKVGEGVGNIANFELDQYLPQGFKTIA 553

Db 403 VSPE-----NITVE-----GEDVKETVT 421

QY 554 SKDYPEVSDYFTVPTSLAYMASQTIFFPHAGDYLRVNPQFVAVPKGTDALVRVE 613

Db 422 AKSDSKTTLD-----FSDLLTKNPSVS-----DRISTN 450

QY 614 FNGNAYLENNYKVGKILPIPKLNGTTRT-----AGNKIPVTFMANAYLDNOSTIVE 667

Db 451 YKNT-----DNHKAIEATIKNLKNSQVTLKAKDSDGNVVEKIF-----T 493

QY 668 VPILKENOTDKPSILPQPKRKAQENSKLDEK---EETPKSEKVEKELSETGN----- 720

Db 494 ITVQKEER-----QVPKTPEDKSDTEKVPQEPKSNQKQLEKSAQOEKLE 545

QY 721 STSNSTLEVPVY-----DPVQEKVAF-----AESYGHKL 751

Db 546 ERAIKELMEQPEIPSNPEYGIOKSIWESQKEPIQEAITSFKKIIGDSSSKYTYEHPKY 605

QY 752 ENVLPMNDGTIELYPSGVEIKKNMADFTGEAPOGNGENKPSNGKVS-----GTVENQ 806

Db 606 KSHFANYQLHAQM-----EMLTRKVVQYMKYIPDPAEIKKIFESDMKRTKEDNNGSLEND 660

QY 807 PTE-----NK-----PADSLPEADNEKVPKPNSTDNGLMNLNPEGNVGSDPMLD 849

Db 661 ALKGYFERKYLTPFKIKQIVDDFDKVEQOQAPPI-PENSE-----MD 703

QY 850 PALEA-----PAVDPVQOEKLEKFTASYGLGL-----DSVIFNM 884

Db 704 QAKERAKIAVSKYMSKVLGDGVHQHLOKKNHSKIVDLFKELEBAIKQOTIFDID 755

RESULT 14

US-10-369-493-22154

Sequence 22154, Application US/10369493

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 22154

LENGTH: 1478

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-10-369-493-22154

Query Match 4.28; Score 197; DB 6; Length 1478;

Best Local Similarity 17.79; Pred. No. 2.2e-05;

Matches 183; Conservative 139; Mismatches 307; Indels 404; Gaps 47;

QY 34 KDSLSEAEAAAAQAYAKEKGLTPPSTD-----HODSGNTEAKGAEIYNRV-----KAA 82

Db 251 QEDISNSTSESALSPTKS-GPSKTDKELHSTSTHQTKSASSLYRRSFISLRSS 309

QY 83 KKVLDPMYNOYVEVKNGLI-----IPSYD-----HYHNKIFE----- 119

Db 310 SNASAKSPSNIKLSIPARPHSIIESNTLTKSPASPSPSPSIFRRHKSSESLL 369

QY 120 --WEDEGLVEA-----PRGYSLEDLATVYVYVEPRNASDHVRKNKADQSKDDEK 171

Db 370 NSLFGSGIGEAEPTKPNQGHSL-----SSENLAGK-----SKHYETNVS 410

QY 172 DEVSEPTHPESDEKEN--HAGLNPSADNLYKPSDTTETETEEAEEDTDBAEIPGTPSI 229

Db 411 SPLKQSSLTSDDKGNLWNRKFRKSKQIGVSPNTVAVYTSQE-----TPSLKS 458

QY 230 NA-----METL-----TGLKSSLLGTRKDNNTISAEDVSLALLKESQ 267

Db 459 NSSTATLTOTADVNTIPSPSSPPPIPKTANRSLVISTEDTPKIS-----STASFKETY 514

QY 268 PAPIQGPQIGQPLPNNSLATPSPSPINPGTSHEKH-----EEDGYGFANR-- 315

Db 515 PDCINPK-----IVP-----VFNQKYSVRNELLQKRYFLPKTKTGLNDSNKYI 560

QY 316 IIADESGFV-----MSH-----GDSNHYF 335

Db 561 LVTKDNVSFVPLNLKSVAKLSFESALTKLGINHNKNTVHTDFDCDIGAATPDDTLEF 620

QY 336 FKDL-----TEQIKAAOKHLEEVKTSUNG--LDSLSSEHQDYPGNKAKMDLD- 383

Db 621 LKSLFLNTSGKIYKQDKMLQKPKAPLTSNNVPLKSVKSKSMRSGTSSLIASD 680

QY 384 -----KKIEKIAGIMKQY-----GVRESIYVNEK 410

681 VSIPTSSDITSFDEHSGSGRRYPQTPSYYYDRVSNINPTNTEELNWNKE---VLSHEE 737
411 NA--IIPSGDHHADPIDEHK-----PVGIGHSHSNYELFKPEGVA---KREGNKVY 459
738 NAKPVFTKSPKLENLDPKGLNIPITENESKSFVLRKDEGTEDFNHRRSPY 797
460 TGEEL-----TNVNLKNSFNQNFNFIANGQRKVSFS-----FPPLEKLL 502
798 TPTELAPKREAPKPPANTSPORTLSTSK-QNKPRLVRASTKISRKSKPLPPOL---L 853
503 GINMLVKLITPDGKVLKSVGKVGEGVGNIANFELDQPYLPQGTFKYTIASKDYP--- 558
854 SSPIEASSSPD-----SLTSSITPAST--HVLPDPKPGAND 889
559 -----EVSVDGFTVPTSLAYKN-----ASOTIFPPHAGDTYLRVNPQPAVPRG 603
890 VMRLATQDOSTSP-SLKMAQKYNRSNSTVSTNSIFS----- 929
604 TDALVRVDFEFGNAYLENNYKVGKILPDKLNQTTTAGNKIPVTFMANAYLDNQST 663
930 -----PSPLLRGNSKRV---VSSTSAADIFEENDIT 958
664 YIVEVPILEKENTDKPSILPOFKRNKAQENSKLDKVEEPTSEKVEKEKLSGTGNS 723
959 FADAPPWFSDSDSDSS-----SSDDIWSKKKTAPETNNENKDEKSDNSSTHS 1009
724 NSTLEEVPTVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKNMADTGA 783
1010 DEIFYDSQTDKMERKT-----FRPSPEVYQNLKFP---F 1043
784 PQNGENKPSNGKVTGTGVENOPTENKPADSL-----PEANKEPKPKENST 831
1044 PRAN-LDKPITEGAS-----TPSPKSLDLSLSPKNVASSRTEPSTPS-RVPDPDSY 1094
832 D-----NGMLNP 838
1095 EFTQDGLNGNKP 1107

RESULT 15
US-09-797-385-6
Sequence 6, Application US/09797385
GENERAL INFORMATION:
APPLICANT: Tal, Joseph Y.
Blake, Milan S.
TITLE OF INVENTION: Non-Iga Fc Binding Forms of the Group B
Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Keseler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/797,385
FILING DATE: 01-Mar-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/923,992
FILING DATE: 05-SEP-1997
APPLICATION NUMBER: US 60/024,707
FILING DATE: (v11)
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-797-385-6

Query Match 4.1%, Score 195.5; DB 5; Length 1128;
Best Local Similarity 17.3%; Pred. No. 1.8e-05;
Matches 15; Conservative 130; Mismatches 296; Indels 309; Gaps 34;

QY 166 DEDKEHDEVSPTHP-----ESDEKENHAGLNPSADNLKPSDTEETEERAEADTDEAIP 222
DB 7 DSVKTEVAAPKPPSMAQTDOGNSSSELETTTRMEITPTDIKKAVEPVKTAGETSAT 66
QY 223 GTPSIRQNMETLGLKSSL---LLGTKNNTISAEV-----DSSLAL----- 262
DB 67 DTGKREKQLOQWKNLNDVNTILSHEQKNEFKTKIDETNDSALLELENOFNETNRL 126
QY 263 -LKESOPAPIQGPQIGQPTLPN-----NSLATPSPSLPINPQTSHE----- 302
DB 127 HIKQHEEVEKDKAKQOKTLAGOSDTKVDSLNDKELNHOQSPVEKMAEPKGITNEDKDSM 186
QY 303 -----KHEEDGVGFANRIIADESGFVMSGDSN-----HYFFKOLTEQ----- 344
DB 187 LKKIEDIRKQQAQADKEDAEVYREELGKLFSSFKAGLDQEIHEHVKKTSSEENTQKV 246
QY 345 -----IKAAQHLEEVKTSNGLDLSLSEHQ-----DYPGNAKEMKDLKRIEKEI 390
DB 247 DEHYANSQNLAKSLEB-----LDKATNEQATQVKNQFLENAOKLKEQOPLIKETN 299
QY 391 AGIMKQYGVKRESI-----VYKKEKNAIIPSGDHHADPIDHRKPVGIGH 436
DB 300 VKLYKAMSESLQVEKELKHSEANLEDLVAKSKEIVREYEGKLNQKLPKOLE-EE 358
QY 437 SHSNYELFKPEGVAKEGKNTVTGEELTNVYVLLKNSTFNNOFTLANGOKRVFSFPP 496
DB 359 AHSKL-----KQVVEDEFKFKTKTSEQVTPKKRVKRDLAANE-----NNOQKIELTSP 406
QY 497 ELEKGLGINMLVKLITPDGKVLKSVGKVGEGVGNIANFELDQPYLPQGTFKYTIASKD 556
DB 407 E-----NITVYE-----GEDVKFTVTAKS 425
QY 557 YPEVSDGFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQPAVPRGTDALVRVDFEFG 616
DB 426 DSKTTLD-----FSDLLTKYNPSVS-----DRISTNTYKT 454
QY 617 NAVLENNYKVGKILPDKLNQTTTRT-----AGHKIPVTFMANAYLDNQSTYIVEVPI 670
DB 455 NT---DNHKLAEITIKNLKLNESQVTTLAKDDSGNVKFTP-----TIV 497
QY 671 LEKENQTKPSILPOFKRNKAQENSKLDKVEEPTSEKVEKEKLSGTG-----STS 723
DB 498 QKKEE-----QVFKTPEQKDSKTEEVKQPEKPSKNDKQLOELIKRQAQOQLEKLEKA 549
QY 724 NSTLEEVPTV-----DPVQEKVAF-----AESYGMKLENV 754
DB 550 IKLEMEQPEIPSNEPYGIQKSIWESQKEIQEITATFKKLIIGDSSSKYTYEYFNKYSK 609
QY 755 LFNDGTFIELYLPDSGEVKKKNMADFTGEAPQNGENKPSNGKVS-----GVENQOPT 809
DB 610 FMYQLHAQN-----EMLTRKVVQYNNKYPDNAEIKKIFESDMKRTKEDNYGSLNDALK 664
QY 810 -----NK-----PADSLPEAPNEKPKVPENSTNGMLNPEGNVGSDFMLDPAL 852
DB 665 GYFEKLYTFPNKIKQIVDDDKKVEQDQAPPI-PENSE-----MDQAK 707
QY 853 EEA-----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 884

9:21:11 2003

EKAKI AVSKYMSKVLDGVHQHQLQKNHSHKI VDLFKELEAI KQOTIFDID 756

5. _ch completed: March 27, 2003, 12:04:41
Job time : 56 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:08 ; Search time 43 Seconds
(without alignments)
2807.560 Million cell updates/sec

Title: US-09-884-465A-332
Perfect score: 4728
Sequence: 1 MQTYTDETOVAKLAKYK.....IELRLPSGEVIKKNLSDFIA 906

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4728	100.0	907	23	AAU84065 S. pneumoniae deri
2	4723	99.9	907	23	AAU84061 S. pneumoniae deri
3	4705	99.5	907	23	AAU84063 S. pneumoniae deri
4	4675	98.9	901	23	AAU84067 S. pneumoniae deri
5	4672	98.8	901	23	AAU84068 S. pneumoniae deri
6	4663	98.6	901	23	AAU84072 S. pneumoniae deri
7	4663	98.6	901	23	AAU84074 S. pneumoniae deri
8	4656.5	98.5	902	23	AAU84076 S. pneumoniae deri
9	4652.5	98.4	902	23	AAU84078 S. pneumoniae deri
10	4624	97.8	895	23	AAU84069 S. pneumoniae deri

11	4619	97.7	895	23	AAU84070 S. pneumoniae deri
12	4615	97.6	895	23	AAU84073 S. pneumoniae deri
13	4615	97.6	895	23	AAU84075 S. pneumoniae deri
14	4608.5	97.5	896	23	AAU84077 S. pneumoniae deri
15	4608.5	97.5	896	23	AAU84080 S. pneumoniae deri
16	4604.5	97.4	896	23	AAU84079 S. pneumoniae deri
17	4571	96.7	889	23	AAU84071 S. pneumoniae deri
18	4560.5	96.5	890	23	AAU84081 S. pneumoniae deri
19	4518	95.6	1152	23	AAU84054 S. pneumoniae deri
20	4518	95.6	1378	23	AAU84053 S. pneumoniae deri
21	4229.5	89.5	913	23	AAU84059 S. pneumoniae deri
22	4229.5	89.5	1139	23	AAU84055 S. pneumoniae deri
23	4147	87.7	1365	23	AAU84057 S. pneumoniae deri
24	4014.5	84.9	1238	23	AAU84056 S. pneumoniae deri
25	3848.5	81.4	1136	23	AAU84058 S. pneumoniae deri
26	3596	76.1	999	23	AAU84051 S. pneumoniae deri
27	3375	71.4	1019	21	AAU84051 Streptococcus pneu
28	3375	71.4	1019	21	AAU84051 Streptococcus pneu
29	3375	71.4	1019	21	AAU84051 Streptococcus pneu
30	3375	71.4	1039	21	AAU84021 Streptococcus pneu
31	3375	71.4	1039	21	AAU84021 Streptococcus pneu
32	3359	71.0	1019	21	AAU84021 Streptococcus pneu
33	3359	71.0	1019	21	AAU84021 Streptococcus pneu
34	3359	71.0	1019	21	AAU84021 Streptococcus pneu
35	3357	71.0	1019	21	AAU84021 Streptococcus pneu
36	3356	71.0	1019	21	AAU84021 Streptococcus pneu
37	3300.5	69.8	807	23	AAU83997 Streptococcus pneu
38	3298	69.8	632	23	AAU84009 Streptococcus pneu
39	3298	69.8	901	23	AAU84056 Streptococcus pneu
40	3298	69.8	907	23	AAU84060 Streptococcus pneu
41	3298	69.8	907	23	AAU84084 Streptococcus pneu
42	3293	69.6	632	23	AAU84006 Streptococcus pneu
43	3293	69.6	632	23	AAU84013 Streptococcus pneu
44	3288	69.5	632	23	AAU83998 Streptococcus pneu
45	3288	69.5	644	23	AAU83996 Streptococcus pneu

ALIGNMENTS

RESULT 1
AAU84065
ID AAU84065 standard; Peptide; 907 AA.
XX
AC AAU84065;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, VP94.
XX
DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; muten;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating

enting streptococcal infections such as otitis media, meningitis, and bacteraemia

Claim 1; Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 907 AA:

Query Match	100.0%	Score 4728;	DB 23;	Length 907;
Best Local Similarity	100.0%;	Pred. No. 6.1e-275;		
Matches 906;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MQITYTDEIQVAKLAGKTYTDEGYIFDTSWIKKOSLSAEARAAQAYAKERGLTPPSTD	60		
DB 2	MQITYTDEIQVAKLAGKTYTDEGYIFDTSWIKKOSLSAEARAAQAYAKERGLTPPSTD	61		
QY 61	QDSGTEAKGAIAIYRKAANKVPLDRMPYNIQYTVFVNGSLIIPSYDHYHNKFEW	120		
DB 62	QDSGTEAKGAIAIYRKAANKVPLDRMPYNIQYTVFVNGSLIIPSYDHYHNKFEW	121		
QY 121	FDEGLYAPKGYSLIEDLATVYVPRNASDHRVKNKADQSDKDEHDEVSSEPTHP	180		
DB 122	FDEGLYAPKGYSLIEDLATVYVPRNASDHRVKNKADQSDKDEHDEVSSEPTHP	181		
QY 181	ESDEKENHAGLNPSAOLYKPSDTTETEEAEDTTDEAIEPTPSIRONAMETLTGLKS	240		
DB 182	ESDEKENHAGLNPSAOLYKPSDTTETEEAEDTTDEAIEPTPSIRONAMETLTGLKS	241		
QY 241	SLLGTTKDNNTISAEVDSLLALKESQAPIQGPTLPNNSLATPSPSLPINFPTS	300		
DB 242	SLLGTTKDNNTISAEVDSLLALKESQAPIQGPTLPNNSLATPSPSLPINFPTS	301		
QY 301	HEKHEEDGYGFDANRIIAEDSGFVMSHGSDSNHYFFKKDLTEQIKAAQKHLEEVKTSN	360		
DB 302	HEKHEEDGYGFDANRIIAEDSGFVMSHGSDSNHYFFKKDLTEQIKAAQKHLEEVKTSN	361		
QY 361	GLDSLSEHQDYPGNAREMDLKKIEKTAGIMQYGVKRESIVYNKKNALIIYPSGDH	420		
DB 362	GLDSLSEHQDYPGNAREMDLKKIEKTAGIMQYGVKRESIVYNKKNALIIYPSGDH	421		
QY 421	HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEEITNVNLLKNSTFNQ	480		
DB 422	HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEEITNVNLLKNSTFNQ	481		
QY 481	FTLANGKRVSFSPPELEKLGIMLVKLITPDGKLVKSVKVGEGVGNANTFELQ	540		
DB 482	FTLANGKRVSFSPPELEKLGIMLVKLITPDGKLVKSVKVGEGVGNANTFELQ	541		
QY 541	PYLPQOTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQITFTFPFHAGDTYLRVNPQFV	600		
DB 542	PYLPQOTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQITFTFPFHAGDTYLRVNPQFV	601		

QY 601	PKGTDALVRVDFEHGNAVLNNYKVGEIKLPIPKLNOQTTRTAGNKIPVTMANAYLDN	660		
DB 602	PKGTDALVRVDFEHGNAVLNNYKVGEIKLPIPKLNOQTTRTAGNKIPVTMANAYLDN	661		
QY 661	QSYIIVVEPILEKENOTDKPSILPQFKRKAQNSKLDEKVEEPTSEKVEKEKLSGTCN	720		
DB 662	QSYIIVVEPILEKENOTDKPSILPQFKRKAQNSKLDEKVEEPTSEKVEKEKLSGTCN	721		
QY 721	STNSSTLEEVPTVDPQVKAKFAESYGKLENVLFNMDGTIELYLPSEGVKKNMAFT	780		
DB 722	STNSSTLEEVPTVDPQVKAKFAESYGKLENVLFNMDGTIELYLPSEGVKKNMAFT	781		
QY 781	GEAPOGNGENKPSGKSTGTVENOPTENKPADSLPEAPNEKPVKPESTONGMLNPEG	840		
DB 782	GEAPOGNGENKPSGKSTGTVENOPTENKPADSLPEAPNEKPVKPESTONGMLNPEG	841		
QY 841	NVGSDDPMLDPALPEAPVDPQVKLEKFTASYGLGDSVIFNMDGTIELYLPSEGVKKN	900		
DB 842	NVGSDDPMLDPALPEAPVDPQVKLEKFTASYGLGDSVIFNMDGTIELYLPSEGVKKN	901		
QY 901	LSDFIA 906			
DB 902	LSDFIA 907			

RESULT 2

AAU84061

ID: AAU84061 standard; Peptide; 907 AA.

XX AC AAU84061;

DT 08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, VP90.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutagen;
KW BVH-11-2.

XX Streptococcus pneumoniae.
OS Synthetic.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (iii) encoding (i) is useful in DNA immunisation techniques. The streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (iii) is useful for designing DNA probes for use in detecting the presence of streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Query Match 99.9%; Score 4723; DB 23; Length 907;
Best Local Similarity 99.9%; Pred. No. 1.2e-274;
Matches 905; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOITTTDEIOVAKLAGYTTEDGYIFTSTWIKKDSLEAERAAQAYAKEKGLTPPSTD 60
DB 2 MOITTTDEIOVAKLAGYTTEDGYIFTSTWIKKDSLEAERAAQAYAKEKGLTPPSTD 61
QY 61 HODSGNTEAKGAEATYNNRKAAYKVPLOMPYNTQYTVVEVNGSLIIPSYDHYHNKEFW 120
DB 62 HODSGNTEAKGAEATYNNRKAAYKVPLOMPYNTQYTVVEVNGSLIIPSYDHYHNKEFW 121
QY 121 FDEGLYEPKGYSLIEDLATVYVYVEPRNASHVRKNKADQSKPDEKDEHVESEPH 180
DB 122 FDEGLYEPKGYSLIEDLATVYVYVEPRNASHVRKNKADQSKPDEKDEHVESEPH 181
QY 181 ESDEKENHAGLNPADNLYKPKSTDETEEEAEDTDEAIPGTPSIRONAMETITGLKS 240
DB 182 ESDEKENHAGLNPADNLYKPKSTDETEEEAEDTDEAIPGTPSIRONAMETITGLKS 241
QY 241 SLLGTFKDNNTISAEDVSLALLKESQAPIQGQPTLPNNSLATPSPSLNPCTS 300
DB 242 SLLGTFKDNNTISAEDVSLALLKESQAPIQGQPTLPNNSLATPSPSLNPCTS 301
QY 301 HEKHEEDGFGFANRIIAEDSGFVNSHSDSNHYFFPKDLTEQIKAAQKHLEEVKTSN 360
DB 302 HEKHEEDGFGFANRIIAEDSGFVNSHSDSNHYFFPKDLTEQIKAAQKHLEEVKTSN 361
QY 361 GLDLSLSEHDYPCGNAKEMKDKLIEEKIAGIMKQYGVKRSIVVVKKNALIIYPSGDH 420
DB 362 GLDLSLSEHDYPCGNAKEMKDKLIEEKIAGIMKQYGVKRSIVVVKKNALIIYPSGDH 421
QY 421 HHADPTDEHKPVGIGSHSNYELFKPEEGVAKKEGKNVYTGELTNVLLKNSFTNNQN 480
DB 422 HHADPTDEHKPVGIGSHSNYELFKPEEGVAKKEGKNVYTGELTNVLLKNSFTNNQN 481
QY 481 FTLANGKRVSFPPPELEKLGINMLVLITPDGKVLKSVKGVGEGVGNIANFELDQ 540
DB 482 FTLANGKRVSFPPPELEKLGINMLVLITPDGKVLKSVKGVGEGVGNIANFELDQ 541
QY 541 PYPGQFTKTIASKDYPSYSGTFTVPSLAYKASQTIFFPHAGTYLVRNQFAV 600
DB 542 PYPGQFTKTIASKDYPSYSGTFTVPSLAYKASQTIFFPHAGTYLVRNQFAV 601
QY 601 PKGTDLVRVDFDEPHGNALYNNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLON 660
DB 602 PKGTDLVRVDFDEPHGNALYNNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLON 661
QY 661 QSTYIVVEPILKENQTDKPSILPQFRKNAQENSKLDEKVEPKTSSEKVEKELSETGN 720
DB 662 QSTYIVVEPILKENQTDKPSILPQFRKNAQENSKLDEKVEPKTSSEKVEKELSETGN 721
QY 721 STNSLTLEEVTPDQVQEKVAKFAESGKMLNVLFNMDGTIELYLSGGEVVKKNMADFT 780
DB 722 STNSLTLEEVTPDQVQEKVAKFAESGKMLNVLFNMDGTIELYLSGGEVVKKNMADFT 781
QY 781 GEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTNDGMLNPEG 840

DB 782 GEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTNDGMLNPEG 841
QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDCGTIELRLPSPGVIKKN 900
DB 842 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDCGTIELRLPSPGVIKKN 901
QY 901 LSDFIA 906
DB 902 LSDFIA 907

RESULT 3
AAU84063
ID AAU84063 standard; Peptide; 907 AA.
XX
AC AAU84063;
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, VP92.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WC200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA0908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
FI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Claim 1; Page -: 113pp; English.
XX
CC The invention describes an isolated polypeptide (i) with 70-90A
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11 or chimeric sequences derived from them. A vaccine (ii)
CC comprising (i) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (ii) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (iii) encoding (i) is useful in DNA immunisation
CC techniques. The streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (iii) is useful for designing DNA
CC probes for use in detecting the presence of streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX

AAU84067 standard; Peptide; 901 AA.		ID	XX
AAU84067;		AC	XX
08-MAY-2002 (first entry)		DT	XX
S. pneumoniae derived chimeric peptide, VP109.		XX	XX
BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;		DE	XX
pneumonia; streptococcal bacterial infection; mutant; mutein;		KW	XX
BVH-11-2.		KW	XX
Streptococcus pneumoniae.		XX	XX
Synthetic.		OS	XX
WO200198334-A2.		XX	XX
27-DEC-2001.		PN	XX
19-JUN-2001; 2001WO-CA00908.		XX	XX
20-JUN-2000; 2000US-212683P.		PD	XX
(SHIR-) SHIRE BIOCHEM INC.		XX	XX
Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;		XX	XX
WPI; 2002-12272/16.		XX	XX
New Streptococcus pneumoniae BVH-3 and BVH-11 variant and		XX	XX
epitope-bearing polypeptides, useful as vaccine components for treating		XX	XX
or preventing streptococcal infections such as otitis media,		XX	XX
meningitis, and bacteraemia		XX	XX
Claim 1; Page -: 113pp; English.		XX	XX
The invention describes an isolated polypeptide (I) with 70-90%		XX	XX
identity to streptococcus pneumonia protein BVH-3, BVH-11, variants of		XX	XX
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)		XX	XX
comprising (I) is useful for therapeutic or prophylactic treatment of		XX	XX
meningitis, otitis media, bacteraemia or pneumonia infection in an		XX	XX
individual susceptible to these disorders. (II) is also useful for		XX	XX
therapeutic or prophylactic treatment of any streptococcal bacterial		XX	XX
infection (e.g., caused by Streptococcus pneumoniae, group A		XX	XX
Streptococcus such as Streptococcus pyogenes, group B Streptococcus or		XX	XX
Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or		XX	XX
Staphylococcus aureus) in an individual susceptible to the infection.		XX	XX
A polynucleotide (III) encoding (I) is useful in DNA immunisation		XX	XX
techniques. The Streptococcus polypeptides are useful in a diagnostic		XX	XX
test for S. pneumoniae infection. (III) is useful for designing DNA		XX	XX
probes for use in detecting the presence of Streptococcus in a biological		XX	XX
sample suspected of containing the bacteria. The DNA probes may also be		XX	XX
used for detecting circulating S. pneumonia nucleic acid in a sample for		XX	XX
diagnosing streptococcal infections. This sequence represents a chimeric		XX	XX
gene created from fragments and variant fragments of Streptococcus		XX	XX
pneumoniae genes, described in the method of the invention.		XX	XX
Note: This sequence does not appear in the specification but has		XX	XX
been created according to information given in the invention.		XX	XX
Sequence 901 AA:		XX	XX
Query Match 98.9%; Score 4675; DB: 23; Length 901;		XX	XX
Best Local Similarity 99.3%; Pred. No. 9, 1e-272;		XX	XX
Matches 900; Conservative 0; Mismatches 0; Indels 6; Gaps 1;		XX	XX
QY 1 MQITTTDDDEIQVAKLAGKTYTDEGTYFTSWIKKDSLEAERAAQAAYAKKGLTPPSTD 60		QY	1
DB 2 MQITTTDDDEIQVAKLAGKTYTDEGTYFTSWIKKDSLEAERAAQAAYAKKGLTPPSTD 61		DB	2
QY 61 HQDSGTEAKGAEATYNRVKAAYKVPDRMPYNYQTYVEVNGSLIIPSYDHYHNKFEW 120		QY	61
DB 62 HQDSGTEAKGAEATYNRVKAAYKVPDRMPYNYQTYVEVNGSLIIPSYDHYHNKFEW 120		DB	62
QY 121 FDEGLYEAAPKGSLEDLATVATYVYEPNASHVRKADQSKPDEKHEVSEPH 180		QY	121
DB 122 FDEGLYEAAPKGSLEDLATVATYVYEPNASHVRKADQSKPDEKHEVSEPH 181		DB	122
QY 181 ESDKENHAGLAPSDNLYKPTDTEETEEAEADTDEAETPGTPIRQNAWETLTGLKS 240		QY	181
DB 182 ESDKENHAGLAPSDNLYKPTDTEETEEAEADTDEAETPGTPIRQNAWETLTGLKS 241		DB	182
QY 241 SLLGTGKDNNTISAEDVSLALLKESQAPIQGPIQGTPLPNNSLATPSPSLINPGTS 300		QY	241
DB 242 SLLGTGKDNNTISAEDVSLALLKESQAPIQGPIQGTPLPNNSLATPSPSLINPGTS 301		DB	242
QY 301 HEKHEEDGYGDFDANRIIAEDSGFVMSHGDSNHYFFKKDLTTEQIKAAQKHLEEVKTSN 360		QY	301
DB 302 HEKHEEDGYGDFDANRIIAEDSGFVMSHGDSNHYFFKKDLTTEQIKAAQKHLEEVKTSN 361		DB	302
QY 361 GLDLSHEDQYDGNKAKEMDKLTKTEKIAGIMKQYGVKRESIVVNEKNAIIPSGDH 420		QY	361
DB 362 GLDLSHEDQYDGNKAKEMDKLTKTEKIAGIMKQYGVKRESIVVNEKNAIIPSGDH 421		DB	362
QY 421 HHADPIDEHRPVGIGHSHNYELFKPEGVAKKEGKNTYTGELTNVNLKNSTFNQ 480		QY	421
DB 422 HHADPIDEHRPVGIGHSHNYELFKPEGVAKKEGKNTYTGELTNVNLKNSTFNQ 481		DB	422
QY 481 FTLANGKRVSPFPPELEKGLINMLVLTDPGKVLKSVKGVGEGVGNIANFELQ 540		QY	481
DB 482 FTLANGKRVSPFPPELEKGLINMLVLTDPGKVLKSVKGVGEGVGNIANFELQ 541		DB	482
QY 541 PYPGQTFKTYTISKDYPEVSYDGTFTVPTSLAYKMASQTFYFPFHAGDTYLRVNPQFV 600		QY	541
DB 542 PYPGQTFKTYTISKDYPEVSYDGTFTVPTSLAYKMASQTFYFPFHAGDTYLRVNPQFV 601		DB	542
QY 601 PKGTDLVRVDFDEPHGNAYLNNYKVEIKLPIPKLNOGTTRAGNKIPVTFMANAYLON 660		QY	601
DB 602 PKGTDLVRVDFDEPHGNAYLNNYKVEIKLPIPKLNOGTTRAGNKIPVTFMANAYLON 661		DB	602
QY 661 QSTVIVPEVPILEKENQDKPSILPQPKRKAQNSKLDEKVEEPTKSEKVEKLSGTGN 720		QY	661
DB 662 QSTVIVPEVPILEKENQDKPSILPQPKRKAQNSKLDEKVEEPTKSEKVEKLSGTGN 721		DB	662
QY 721 STNSTLEEVTVPDQVEKVAFAESTGMKLENVFNQGTIELIYLPSPGVIKKNADFT 780		QY	721
DB 722 STNSTLEEVTVPDQVEKVAFAESTGMKLENVFNQGTIELIYLPSPGVIKKNADFT 781		DB	722
QY 781 GEAPGNGENKPSNGKVTSTGTVENQPTENKPADSLPEAPNEKPVKPDENSTDMNLP 840		QY	781
DB 782 GEAPGNGENKPSNGKVTSTGTVENQPTENKPADSLPEAPNEKPVKPDENSTDMNLP 841		DB	782
QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDDGTIELRLPSGEVIKKN 900		QY	841
DB 842 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDDGTIELRLPSGEVIKKN 901		DB	842
QY 901 LSDFTA 906		QY	901
DB 902 LSDFTA 907		DB	902
RESULT 4		XX	XX
AAU84067		XX	XX

Fri Mar 28 09:21:09 2003

us-09-884-465a-332.rag

121 FDEGLYEPKGYSLIEDLATVYKYYVPRNASDHVRNKADQDQSKPDEDKEHDEVSEPTHP 180
116 FDEGLYEPKGYSLIEDLATVYKYYVPRNASDHVRNKADQDQSKPDEDKEHDEVSEPTHP 175
181 ESDEKENHAGLNPSADNLYKPSYTDTEETEAEEDTDEAIPOTPSIRQNAHMETLGLKS 240
176 ESDEKENHAGLNPSADNLYKPSYTDTEETEAEEDTDEAIPOTPSIRQNAHMETLGLKS 235
241 SLLGTDKNTNTISAEVDSLLALKESQAPIQGPIQGPQLPNNSLATPSPSLPINPGTS 300
236 SLLGTDKNTNTISAEVDSLLALKESQAPIQGPIQGPQLPNNSLATPSPSLPINPGTS 295
301 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 360
296 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 355
361 GLDLSSEHQDYPGNAKEMDKLIEKTAGIMKQYGVKRESIVVKNENAIYPSGDH 420
356 GLDLSSEHQDYPGNAKEMDKLIEKTAGIMKQYGVKRESIVVKNENAIYPSGDH 415
421 HADPIDHKKPVGIGHSHSNYELEKPEGEVAKKEGNKYVTGEELTNVNLKSTFNQON 480
416 HADPIDHKKPVGIGHSHSNYELEKPEGEVAKKEGNKYVTGEELTNVNLKSTFNQON 475
481 FTLANGQRVSFSFPPELEKLGIMLVKLITPDGKVLKESGVKFGVGVGNIANFELDQ 540
476 FTLANGQRVSFSFPPELEKLGIMLVKLITPDGKVLKESGVKFGVGVGNIANFELDQ 535
541 PYLPQGTFTYTTASKDYPEVSDGFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 600
536 PYLPQGTFTYTTASKDYPEVSDGFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 595
601 PRGTDALRVDFEFGHGNLYENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDN 660
596 PRGTDALRVDFEFGHGNLYENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDN 655
661 QSYTYVEVILEKENQDPSILPQFARNKAGENSCLDEKVEEPTSEKVEKLSGTGN 720
656 QSYTYVEVILEKENQDPSILPQFARNKAGENSCLDEKVEEPTSEKVEKLSGTGN 715
721 STNSNLEEVPTVDPQEVKAFSAESYGMKLENVLFNMDGTIELYLPSEGVKKMAOFT 780
716 STNSNLEEVPTVDPQEVKAFSAESYGMKLENVLFNMDGTIELYLPSEGVKKMAOFT 775
781 GEAPQNGENKPSNGKSVGTGTENQTPENKPADSLPEAPNEKPVKPNSTNGMLNPEG 840
776 GEAPQNGENKPSNGKSVGTGTENQTPENKPADSLPEAPNEKPVKPNSTNGMLNPEG 835
841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRPLPSGEVKK 900
836 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRPLPSGEVKK 895
901 LSDPIA 906
896 LSDPIA 901
RESULT 5
AAU84068
ID AAU84068 standard; Peptide; 901 AA.
XX
XX
AC AAU84068;
XX
DT 08-MAY-2002 (first entry)
XX
XX S. pneumoniae derived chimeric peptide, VP110.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteriaemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX Streptococcus pneumoniae.
OS Synthetic.

XX W0200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Claim 1; Page -; 113pp; English.
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
XX Staphylococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in a diagnostic
XX techniques. The Streptococcus polypeptides are useful for designing DNA
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a chimeric
XX gene created from fragments and variant fragments of Streptococcus
XX pneumoniae genes, described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 901 AA;
XX
XX Query Match 98.8%; Score 4672; DB 23; Length 901;
XX Best Local Similarity 99.2%; Pred. No. 1.4e-271;
XX Matches 899; Conservative 0; Mismatches 1; Indels 6; Gaps 1;
QY 1 MOITYTDDDEIOVAKLAGYTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPPSTD 60
DB 2 MOITYTDDDEIOVAKLAGYTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPPSTD 61
QY 61 HDSCNTEAKGAEIYNKVAKKVPLDRMPYNLQYTYVEVNGSLIIPSYDHYHNKFEW 120
DB 62 HDSCNTEAKGAEIYNKVAKKVPLDRMPYNLQYTYVEVNGSLIIPSYDHYHNKFEW 121
QY 121 FDEGLYEPKGYSLIEDLATVYKYYVPRNASDHVRNKADQDQSKPDEDKEHDEVSEPTHP 180
DB 122 FDEGLYEPKGYSLIEDLATVYKYYVPRNASDHVRNKADQDQSKPDEDKEHDEVSEPTHP 181
QY 181 ESDEKENHAGLNPSADNLYKPSYTDTEETEAEEDTDEAIPOTPSIRQNAHMETLGLKS 240
DB 182 ESDEKENHAGLNPSADNLYKPSYTDTEETEAEEDTDEAIPOTPSIRQNAHMETLGLKS 241
QY 241 SLLGTDKNTNTISAEVDSLLALKESQAPIQGPIQGPQLPNNSLATPSPSLPINPGTS 300
DB 242 SLLGTDKNTNTISAEVDSLLALKESQAPIQGPIQGPQLPNNSLATPSPSLPINPGTS 301
QY 301 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 360
DB 302 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 360

XX PF New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Claim 1; Page -; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX SQ Sequence 901 AA;
 Query Match 98.6%; Score 4663; DB 23; Length 901;
 Best Local Similarity 99.1%; Pred. No. 4.8e-271;
 Matches 896; Conservative 0; Mismatches 2; Indels 6; Gaps 1;
 QY 1 MQITYTDEIOVAKLAGYTTEDGYIFDTSWIKKDSLEAERAAQAAYAKEGLTPSTD 60
 DB 2 MQITYTDEIOVAKLAGYTTEDGYIFDTSWIKKDSLEAERAAQAAYAKEGLTPSTD 61
 QY 61 HQDSGNTAEKAEALYNRVKAACKVPLDRPYNLYQYVEVKNGLIIPSDVHYNIKFEW 120
 DB 62 HQDSGNTAEKAEALYNRVKAACKVPLDRPYNLYQYVEVKNGLIIPSDVHYNIKFEW 121
 QY 121 FDEGLYEAPKGYSLDGLATYKYVPEPRNASDHYRKNKAQDQSKPDEKHEVSETHP 180
 DB 122 FDEGLYEAPKGYSLDGLATYKYVPEPRNASDHYRKNKAQDQSKPDEKHEVSETHP 181
 QY 181 ESDKENHAGLNPADNLYKPSDTETEETEEAEEDTDEAETPGTSPTRONAMETLTGLK 240
 DB 182 ESDKENHAGLNPADNLYKPSDTETEETEEAEEDTDEAETPGTSPTRONAMETLTGLK 241
 QY 241 SLLGTDKNNTISAEVDSILLALKESQAPIQGPIQIGOPTLPNNSLATPSPSLPINFPTS 300
 DB 242 SLLGTDKNNTISAEVDSILLALKESQAPIQGPIQIGOPTLPNNSLATPSPSLPINFPTS 301
 QY 301 HEKHEEDGYFDANRIAEDESGFVNSHGSDNHFFKFKDLEEQIKAAQKHEVKTSHN 360
 DB 302 HEKHEEDGYFDANRIAEDESGFVNSHGSDNHFFKFKDLEEQIKAAQKHEVKTSHN 355
 QY 361 GLDLSSSHEDQYPGNKAKEMKOLDKKEIKIAGINKQYGVKRESIVVNEKNAIYPSGDH 420
 DB 356 GLDLSSSHEDQYPGNKAKEMKOLDKKEIKIAGINKQYGVKRESIVVNEKNAIYPSGDH 415
 QY 421 HHADPIDEHKPVGIGHSHSNYELFPPEGVAKKEGKNKYTTGELTNVYVNLKSNFNNQ 480
 DB 416 HHADPIDEHKPVGIGHSHSNYELFPPEGVAKKEGKNKYTTGELTNVYVNLKSNFNNQ 475
 QY 481 FTLANGQKRVSPFPPELEKKGILGINKLVKLTIPDGKVLKSGKVGSGVGNIANFELDQ 540
 DB 476 FTLANGQKRVSPFPPELEKKGILGINKLVKLTIPDGKVLKSGKVGSGVGNIANFELDQ 535

RESULT 6

AAU84072
 ID AAU84072 standard; Peptide; 901 AA.
 XX
 AC AAU84072;

XX 08-MAY-2002 (first entry)
 XX
 DE S. pneumoniae derived chimeric peptide, VPI14.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutin;

XX Streptococcus pneumoniae.
 OS Synthetic.
 XX WO200198334-A2.
 PN
 XX 27-DEC-2001.
 PD
 XX 19-JUN-2001; 2001WO-CA00908.
 PP
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI NPI; 2003-122272/16.
 XX
 DR

541 PYLPQTEKYTIASRDYEVSDGTFVTPTSLAYKMASTQIFYPFHAGDTYLRVNPQFAY 600
 536 PYLPQTEKYTIASRDYEVSDGTFVTPTSLAYKMASTQIFYPFHAGDTYLRVNPQFAY 595
 601 PGKTDALVRVDFEFGNAYLENNYKVGSEIKLPIPKLNGQTTTNTAGNKIPVTFMANAYLDN 660
 596 PGKTDALVRVDFEFGNAYLENNYKVGSEIKLPIPKLNGQTTTNTAGNKIPVTFMANAYLDN 655
 661 QSTYIVVPILEKENQDKPSLTPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGN 720
 656 QSTYIVVPILEKENQDKPSLTPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGN 715
 721 STSNSTLEVPVDPQVKAFASGYGKLENVLFNMDGTTIELYLPSEGVYKKNADFT 780
 716 STSNSTLEVPVDPQVKAFASGYGKLENVLFNMDGTTIELYLPSEGVYKKNADFT 775
 781 GEAPGNGENKPSNGKYSTGVNQPENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEG 840
 776 GEAPGNGENKPSNGKYSTGVNQPENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEG 835
 841 NYGSDPMDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTTIELYLPSEGVYKKN 900
 836 NYGSDPMDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTTIELYLPSEGVYKKN 895
 901 LSDFIA 906
 896 LSDFIA 901
 RESULT 7
 AA084074
 ID AA084074 standard; peptide; 901 AA.
 AC AA084074;
 XX
 DT 08-MAY-2002 (first entry)
 DE S. pneumoniae derived chimeric peptide, VP116.
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutin;
 KW BVH-11-2.
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 XX (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 XX WPI; 2002-12272/16.
 XX
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Claim 1, Page -; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for

therapeutic or prophylactic treatment of any streptococcal bacterial
 infection (e.g., caused by Streptococcus pneumoniae, group A
 Streptococcus, group B Streptococcus pyogenes, group B Streptococcus such
 as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 901 AA;
 Query Match 98.6%; Score 4663; DB 23; Length 901;
 Best Local Similarity 99.1%; Pred. No. 4.8e-271;
 Matches 898; Conservative 0; Mismatches 2; Indels 6; Gaps 1;
 Qy 1 MQITYTDEIOVAKLAGYTTEDGTIFDTSWIKKDSLEAERAAQAAYAKEGLTPPSTD 60
 Db 2 MQITYTDEIOVAKLAGYTTEDGTIFDTSWIKKDSLEAERAAQAAYAKEGLTPPSTD 61
 Qy 61 HODSGNTEAKGAEALYNRVKAAYKVPDRMPYLNQYTYVEVNGSLIIPSYDHYHNKFEW 120
 Db 62 HODSGNTEAKGAEALYNRVKAAYKVPDRMPYLNQYTYVEVNGSLIIPSYDHYHNKFEW 121
 Qy 121 FDEGLYEAPKGYSLLEDLATYKYVYPRNASDHVRKNKADQSKDPDEKHEVSEPHP 180
 Db 122 FDEGLYEAPKGYSLLEDLATYKYVYPRNASDHVRKNKADQSKDPDEKHEVSEPHP 181
 Qy 181 ESDEKENAGLNPADNLYKPSDTEETEAEEDTDEAEIPGTPSPISQNAETITGLKS 240
 Db 182 ESDEKENAGLNPADNLYKPSDTEETEAEEDTDEAEIPGTPSPISQNAETITGLKS 241
 Qy 241 SLLIGTKDNNTISAEVDSILLALKESQAPQIOGPTLPNNSLATPSPSLNPQTS 300
 Db 242 SLLIGTKDNNTISAEVDSILLALKESQAPQIOGPTLPNNSLATPSPSLNPQTS 301
 Qy 301 HEKHEEDYGFANDRIIAEDSGFVNSHGDNSHYFFKKDLTEEQKAAQKHLVEVKTSHN 360
 Db 302 HEKHEEDYGFANDRIIAEDSGFVNS -----YFFKKDLTEEQKAAQKHLVEVKTSHN 355
 Qy 361 GLDLSLSSHEQDYPGNAKEMKDLKKIEKGIAGIMKOYGVKRESIVVNEKKNAILYPSGDH 420
 Db 356 GLDLSLSSHEQDYPGNAKEMKDLKKIEKGIAGIMKOYGVKRESIVVNEKKNAILYPSGDH 415
 Qy 421 HHADPIDEHKPVGIGHSHSNYELFPEEGVAKEGKNVYTGELTNVYNLLKNSTFNQ 480
 Db 416 HHADPIDEHKPVGIGHSHSNYELFPEEGVAKEGKNVYTGELTNVYNLLKNSTFNQ 475
 Qy 481 FTLANGQKRVSFSPPELEKLGIMLVKLITPDGKLVLEKSGVFGEGVGNIAFELDQ 540
 Db 476 FTLANGQKRVSFSPPELEKLGIMLVKLITPDGKLVLEKSGVFGEGVGNIAFELDQ 535
 Qy 541 PYLPQTEKYTIASRDYEVSDGTFVTPTSLAYKMASTQIFYPFHAGDTYLRVNPQFAY 600
 Db 536 PYLPQTEKYTIASRDYEVSDGTFVTPTSLAYKMASTQIFYPFHAGDTYLRVNPQFAY 595
 Qy 601 PGKTDALVRVDFEFGNAYLENNYKVGSEIKLPIPKLNGQTTTNTAGNKIPVTFMANAYLDN 660
 Db 596 PGKTDALVRVDFEFGNAYLENNYKVGSEIKLPIPKLNGQTTTNTAGNKIPVTFMANAYLDN 655
 Qy 661 QSTYIVVPILEKENQDKPSLTPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGN 720
 Db 656 QSTYIVVPILEKENQDKPSLTPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGN 715
 Qy 721 STSNSTLEVPVDPQVKAFASGYGKLENVLFNMDGTTIELYLPSEGVYKKNADFT 780
 Db 721 STSNSTLEVPVDPQVKAFASGYGKLENVLFNMDGTTIELYLPSEGVYKKNADFT 780

CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

Query Match 98.5%; Score 4656.5; DB 23; Length 902;
Best Local Similarity 99.0%; Pred. NO. 1.2e-270;
Matches 898: Conservative 0; Mismatches 2; Indels 7;

Qy 901 LSDFA 906
111111

RESULT 8

AA
AC
AA084076;

g. pneumoniae derived chimeric peptide, vp119.

OS Streptococcus pneumoniae.

PN WO200198334-A2.

FD 37-DEC-2001.
XX

XX
PB
70-1111-2000. 2000PS-212683D

PA (SHIR-) SHIRE BIOCHEM INC.

XX WPT: 2002-122272/16

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -

CC The invention describes an isolated polypeptide (I) with 70-908

identity to *Streptococcus pneumoniae* protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by *Streptococcus pneumoniae*, group A *Streptococcus* such as *Streptococcus pyogenes*, group B *Streptococcus* such as *Streptococcus agalactiae*, *S. dysgalactiae*, *S. uberis*, *S. noxia* or *Streptococcus aureus*) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The *Streptococcus* polypeptides are useful in a diagnostic test for *S. pneumoniae* infection. (III) is useful for designing DNA probes for use in detecting the presence of *Streptococcus* in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating *S. pneumoniae* nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of *Streptococcus pneumoniae* genes, described in the method of the invention.

RESULT 9

AAU84078
ID AAU84078 standard; Peptide; 902 AA.
XX
AC AAU84078;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, VP121.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Claim 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
XX Staphylococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a chimeric
XX gene created from fragments and variant fragments of Streptococcus
XX pneumoniae genes, described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 902 AA;

Query Match 98.4%; Score 4652.5; DB 23; Length 902;
Best Local Similarity 99.0%; Pred. No. 28-270;
Matches 898; Conservative 0; Mismatches 2; Indels 7; Gaps 2;

QY 1 MQITYTDEIQVAKLAGKTTEDGYIFDTSWIKKSLSEARAQAAYAKKGLTPPSTD 60
DB 2 MQITYTDEIQVAKLAGKTTEDGYIFDTSWIKKSLSEARAQAAYAKKGLTPPSTD 61
QY 61 HODSGNTEARGAEIYNRVKAAKVPDLRMPYNLOQTYVEKNGSLIIPSYDHYHNKFEW 120

DB 62 HQDSGNTARGAEIYNRVKAAKVPDLRMPYNLOQTYVEKNGSLIIPSYDHYHNKFEW 121
QY 121 FDEGLYEAPKGYSLDGLATVYKYVEPRNASHVRNKAQDQSKPDDEKHEHVSPTHP 180
DB 122 FDEGLYEAPKGYSLDGLATVYKYVEPRNASHVRNKAQDQSKPDDEKHEHVSPTHP 181
QY 181 ESDEKENHAGLNPSADNLYKPSDTDETEEEAEDTDEAEIPGTPSIRONAMETLGLKS 240
DB 182 ESDEKENHAGLNPSADNLYKPSDTDETEEEAEDTDEAEIPGTPSIRONAMETLGLKS 241
QY 241 SLLLTGDKNTISAEDVSLALLESQAPAIQ-GPQIGOPTLPNNSLATPSPSLPNPGT 299
DB 242 SLLLTGDKNTISAEDVSLALLESQAPAIQSGPQIGOPTLPNNSLATPSPSLPNPGT 301
QY 300 SHEKHEEDGYGFDAIRIIAEDSGFVMSHSDSNHYFFKKDLTEEQIAKAAKHLEEVKTS 359
DB 302 SHEKHEEDGYGFDAIRIIAEDSGFVMS-----YFFKKDLTEEQIAKAAKHLEEVKTS 355
QY 360 NGLDLSLSSHEQDYPGNAKEMKDLKKTEEKIAGIMKYGKRESIVVYVKNKNAIIPSGD 419
DB 356 NGLDLSLSSHEQDYPGNAKEMKDLKKTEEKIAGIMKYGKRESIVVYVKNKNAIIPSGD 415
QY 420 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVTGEELTNVYVLLKNSTFNNQ 479
DB 416 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVTGEELTNVYVLLKNSTFNNQ 475
QY 480 NFTLANGOKRVSPFPPELEKKGILNMLVLITPDGKVLKSVKGVGEGVGNIANFELD 539
DB 476 NFTLANGOKRVSPFPPELEKKGILNMLVLITPDGKVLKSVKGVGEGVGNIANFELD 535
QY 540 QPYLPQGTFFKTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFA 599
DB 536 QPYLPQGTFFKTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFA 595
QY 600 VPKGTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLD 659
DB 596 VPKGTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLD 655
QY 660 NQSYIYVEVPILEKENTQDKPSILPOPKRNAQENSKLDEKVEPKTSEKVEKELSETG 719
DB 656 NQSYIYVEVPILEKENTQDKPSILPOPKRNAQENSKLDEKVEPKTSEKVEKELSETG 715
QY 720 NSTNSTLEEYPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNADE 779
DB 716 NSTNSTLEEYPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNADE 775
QY 780 TGEAPQNGENKPSNGKSVGTGTVENQTEKPKADSLPEAPNEKPVKPESTNDGMLNPE 839
DB 776 TGEAPQNGENKPSNGKSVGTGTVENQTEKPKADSLPEAPNEKPVKPESTNDGMLNPE 835
QY 840 GNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKK 899
DB 836 GNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKK 895
QY 900 NLSDFIA 906
DB 896 NLSDFIA 902
RESULT 10
AAU84069
ID AAU84069 standard; Peptide; 895 AA.
XX
AC AAU84069;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, VP11.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain;
KW BVH-11-2.

-ptococcus pneumoniae.
synthetic.

44 PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

XX PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Scaphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX SQ Sequence 895 AA;

Query Match 97.8%; Score 4624; DB 23; Length 895;
Best Local Similarity 98.7%; Pred. No. 1e-268;
Matches 894; Conservative 0; Mismatches 0; Indels 12; Gaps 2;

QY 1 MQITYTDETOQAKLAKYTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPPSTD 60
DB 2 MQITYTDETOQAKLAKYTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPPSTD 61
QY 61 HODSGNTEAKGAEIYNRVAARKVPLDRPYNLYQYVEVKNGLIIPSDYHYNKFEW 120
DB 62 HODSGNTEAKGAEIYNRVAARKVPLDRPYNLYQYVEVKNGLIIPSDYHYNKFEW 121
QY 121 FDEGLYEPKAGYSLEDLATVYKYVEPRNASHVRKKNADQDSKPDDEKHEVSEPTHP 180
DB 122 FDEGLYEPKAGYSLEDLATVYKYVEPRNASHVRKKNADQDSKPDDEKHEVSEPTHP 181
QY 181 ESEKENHAGLNPSADNLYKPSDTETEAEEDTTDEAEIPGTPSRQNAETLTGLKS 240
DB 182 ESEKENHAGLNPSADNLYKPSDTETEAEEDTTDEAEIPGTPSRQNAETLTGLKS 241
QY 241 SLLIGTKDNNTISAEVDSLLALKESOPATQGPQIGOPTLPNNSLATPSPSLINPGTS 300
DB 242 SLLIGTKDNNTISAEVDSLLALKESOPATQGPQIGOPTLPNNSLATPSPSLINPGTS 301

QY 301 HEKHEEDGYGFDANRIIADESGFVMSHGSDNHYFFPKDLTTEQIKAAOKHLEEVKTSN 360
DB 302 HEKHEEDGYGFDANRIIADESGFVMS-----YFFKDLTTEQIKAAOKHLEEVKTSN 365
QY 361 GLDLSLSEHQDYPGNAKEMKDLDKKIEEKIAGIMQYGVKRESIVVNNKKNALIIYPSGDH 420
DB 366 GLDLSLSEHQDYPGNAKEMKDLDKKIEEKIAGIMQYGVKRESIVVNNKKNALIIY- 411
QY 421 HHADPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNLKKNSTNNON 480
DB 412 --ADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNLKKNSTNNON 469
QY 481 FTLANGQKRVSFSPPELEKKGILNMLVLTITPDGKLVLEKVGKVFEGVGNIANFELQ 540
DB 470 FTLANGQKRVSFSPPELEKKGILNMLVLTITPDGKLVLEKVGKVFEGVGNIANFELQ 539
QY 541 PYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYFPHAGDTYLRVNPQFV 600
DB 530 PYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYFPHAGDTYLRVNPQFV 589
QY 601 PKGTDALYRVDFEHGHNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDN 660
DB 590 PKGTDALYRVDFEHGHNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDN 649
QY 661 QSTYIYVEVPILEKENOTDKPSILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSGTGN 720
DB 650 QSTYIYVEVPILEKENOTDKPSILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSGTGN 709
QY 721 STNSSTLEEYPTVDPQVKVAKFAESYGMKLENVLFNMOGTIELYLPSPGEVIRKKNADFT 780
DB 710 STNSSTLEEYPTVDPQVKVAKFAESYGMKLENVLFNMOGTIELYLPSPGEVIRKKNADFT 769
QY 781 GEAPQNGENKPSNGKVSQGTVENOPTENKPADSLPEADNEKPKVPENSTDNGLNPEG 840
DB 770 GEAPQNGENKPSNGKVSQGTVENOPTENKPADSLPEADNEKPKVPENSTDNGLNPEG 839
QY 841 NVGSDPMLDPALEAPAVDPVQVKLEKFTASYGLGDSVIFNMDGTIELRLSPSGEVIKKN 900
DB 830 NVGSDPMLDPALEAPAVDPVQVKLEKFTASYGLGDSVIFNMDGTIELRLSPSGEVIKKN 889
QY 901 LSDFIA 906
DB 890 LSDFIA 895
RESULT 11
AA084070
ID AA084070 standard; Peptide; 895 AA.
XX AC AA084070;
XX AC AA084070;
DT 08-MAY-2002 (first entry)
XX S. pneumoniae derived chimeric peptide, VP112.
DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX OS Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
PN 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
PF 20-JUN-2000; 2000US-212683P.
PR 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX PA

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 XX epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Claim 1; Page -: 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 895 AA;

Query Match 97.7%; Score 4619; DB 23; Length 895;
 Best Local Similarity 98.6%; Pred. No. 2e-368;
 Matches 893; Conservative 0; Mismatches 1; Indels 12; Gaps 2;

QY 1 MQITTTDBIQVAKLAKGTTTGGTIFDTSWIKKDSLSAEARAAQAAYAKKGLPPSTD 60
 DB 2 MQITTTDBIQVAKLAKGTTTGGTIFDTSWIKKDSLSAEARAAQAAYAKKGLPPSTD 61
 QY 61 HDGSGNTEAKGAEAYNRVKAAYKVPDLPYNYLYTVVKNGLSLIPSYDHYHKNPFW 120
 DB 62 HDGSGNTEAKGAEAYNRVKAAYKVPDLPYNYLYTVVKNGLSLIPSYDHYHKNPFW 115
 QY 121 FDEGLYAPKGYSLLEDLATVYKYVEPRNASDVRNKAQDQSKPDEKHEDEVSEPTHP 180
 DB 116 FDEGLYAPKGYSLLEDLATVYKYVEPRNASDVRNKAQDQSKPDEKHEDEVSEPTHP 175
 QY 181 ESDKENHAGLNPADNLYKPTDTEETEEAEADTDEAIPGTSIRONAMETLTGLKS 240
 DB 176 ESDKENHAGLNPADNLYKPTDTEETEEAEADTDEAIPGTSIRONAMETLTGLKS 235
 QY 241 SLLLGTKONTTISAEDVSLALLKESQAPIQGQPTLPNNSLATPSPSLPNTGTS 300
 DB 236 SLLLGTKONTTISAEDVSLALLKESQAPIQGQPTLPNNSLATPSPSLPNTGTS 295
 QY 301 HEKHEDEGDFDANRIIAEDSGFVMSHSDSNHYFFKDLTEQIKAAQKHLEEVKTSN 360
 DB 296 HEKHEDEGDFDANRIIAEDSGFVMSHSDSNHYFFKDLTEQIKAAQKHLEEVKTSN 349
 QY 361 GLDLSLSSHQDYPGNAKEMKDLKIEEKIAGIMQYGVKRESIVVKNKNAIIPSGDH 420
 DB 350 GLDLSLSSHQDYPGNAKEMKDLKIEEKIAGIMQYGVKRESIVVKNKNAIIPSGDH 409
 QY 421 HADPIDEHKPVGIGHSHNYELFKPEEGVAKKGNKVTGELTNVNLKNSTFNPN 480
 DB 410 HADPIDEHKPVGIGHSHNYELFKPEEGVAKKGNKVTGELTNVNLKNSTFNPN 469
 QY 481 FTLANGQKRVSFPPPELEKLGINMLVKLITPDGKLVLEKSVGKVFEGVGNIANFELQ 540

DB 470 FTLANGQKRVSFPPPELEKLGINMLVKLITPDGKLVLEKSVGKVFEGVGNIANFELQ 529
 QY 541 PYPGQTFKTYTIAKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 600
 DB 530 PYPGQTFKTYTIAKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 589
 QY 601 PKGTDALVRVDFEFGHAYLENNYKVGKIKLPIPKLNGQTTAGNKIPVTFMANAYLDN 660
 DB 590 PKGTDALVRVDFEFGHAYLENNYKVGKIKLPIPKLNGQTTAGNKIPVTFMANAYLDN 649
 QY 661 QSTYIVVEPILKEKNDQDKPSILPOFRKKAQENSKLDEKVEEPTSEKVEKELSETGN 720
 DB 650 QSTYIVVEPILKEKNDQDKPSILPOFRKKAQENSKLDEKVEEPTSEKVEKELSETGN 709
 QY 721 STSNSTLEEVTPDVPQVKAFKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNADFT 780
 DB 710 STSNSTLEEVTPDVPQVKAFKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNADFT 769
 QY 781 GEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 840
 DB 770 GEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 829
 QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSVIFNMDGTIELRLPSEGVYKKN 900
 DB 830 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSVIFNMDGTIELRLPSEGVYKKN 889
 QY 901 LSDFTA 906
 DB 890 LSDFTA 895

RESULT 12
 AAU84073
 ID AAU84073 standard; Peptide; 895 AA.
 AC AAU84073;
 DT 08-MAY-2002 (first entry)
 XX S. pneumoniae derived chimeric peptide, VP115.
 DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutain;
 KW BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 OS WO200198334-A2.
 PN 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 PR (SHIR-) SHIRE BIOCHEM INC.
 PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Claim 1; Page -: 113pp; English.
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

(I) is useful for therapeutic or prophylactic treatment of otitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Query Match 97.68; Score 4615; DB 23; Length 895;
 Best Local Similarity 98.68; Pred. No. 3.5e-268;
 Matches 893; Conservative 0; Mismatches 1; Indels 12; Gaps 2;

1 MQIYTTDEIQVAKLAGYTTEDGTYIDTSWIKKDSLEAERAAQAYAKEKGLTPPSTD 60
 2 MQIYTTDEIQVAKLAGYTTEDGTYIDTSWIKKDSLEAERAAQAYAKEKGLTPPSTD 61
 61 HQDSGNTAKGAEIYNRVRAAKVPLDRPYNQYTVVEVNGSLIIPSDYHYNIKFEW 120
 62 HQDSGNTAKGAEIYNRVRAAKVPLDRPYNQYTVVEVNGSLIIPSDYHYNIKFEW 121
 121 FDEGLYAPKGYSLIEDLATVYKYYEPNADSHVRKKAQDSKPDDEKDESEPTHP 180
 122 FDEGLYAPKGYSLIEDLATVYKYYEPNADSHVRKKAQDSKPDDEKDESEPTHP 181
 181 ESKENHAGLNPSADNLYKPTDTEETEEAEDTTDEAETPGTTSIRQNAETLTGLKS 240
 182 ESKENHAGLNPSADNLYKPTDTEETEEAEDTTDEAETPGTTSIRQNAETLTGLKS 241
 241 SLLGTGDNNTISAEVDSLLALKESOPAPITQGPQIGPILPNNSLATPSPSLINFGTS 300
 242 SLLGTGDNNTISAEVDSLLALKESOPAPITQGPQIGPILPNNSLATPSPSLINFGTS 301
 301 HEKHEEDYGFANDRIITAEDSGFVMSHGSDNHYFFKDLTEEOIKAAQKLEEVKTSN 360
 302 HEKHEEDYGFANDRIITAEDSGFVMS-----YFFKDLTEEOIKAAQKLEEVKTSN 355
 361 GLDSSLSEHQDYPGNAKEMKDLKKIEKIAGIMKQYGVKRESIVVAKENAIYPSGDH 420
 356 GLDSSLSEHQDYPGNAKEMKDLKKIEKIAGIMKQYGVKRESIVVAKENAIYYP--- 411
 421 HHADPIDEKHPVGIHSHSNYELFKPEGVAKKGNKYVYTGEEELTNVLLKNSTFNQON 480
 412 --ADPIDEKHPVGIHSHSNYELFKPEGVAKKGNKYVYTGEEELTNVLLKNSTFNQON 469
 481 FTLANGKRVFSFPPELEKLGINMLVKLITPDGKLVKESGVFGVGVGNIAFELDQ 540
 470 FTLANGKRVFSFPPELEKLGINMLVKLITPDGKLVKESGVFGVGVGNIAFELDQ 529
 541 PYLPGQFKTYTISKQDPEVSDGTFTVPTSLAYKMASQTIFFPHFAGDTYLRVNPQFV 600
 530 PYLPGQFKTYTISKQDPEVSDGTFTVPTSLAYKMASQTIFFPHFAGDTYLRVNPQFV 589
 601 PKGTDALVRVDEPHGNAYLENNYKVEIKLPIKLNQGTTRTAGNKIPTFFMANAYLDN 660
 590 PKGTDALVRVDEPHGNAYLENNYKVEIKLPIKLNQGTTRTAGNKIPTFFMANAYLDN 649
 661 QSTYIVVEVPILEKENQDKPSILPQFRNKAQENSKLDKEVEEPTSEKVEKELSETGN 720
 650 QSTYIVVEVPILEKENQDKPSILPQFRNKAQENSKLDKEVEEPTSEKVEKELSETGN 709

QY 721 STNSTLEEVPTDPOEVKAKFAESYGMKLENVLPNMDGTIELYLPSPGVEIKKNMADET 780
 DB 710 STNSTLEEVPTDPOEVKAKFAESYGMKLENVLPNMDGTIELYLPSPGVEIKKNMADET 769
 QY 781 GEAPQNGENKPSGKSTGTSTVENQPTENKPADSLPEAPNEKPVKPNSTDMGMLNPEG 840
 DB 770 GEAPQNGENKPSGKSTGTSTVENQPTENKPADSLPEAPNEKPVKPNSTDMGMLNPEG 829
 QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRPSGEVIKKN 900
 DB 830 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRPSGEVIKKN 889
 QY 901 LSDFIA 906
 DB 890 LSDFIA 895

RESULT 13
 AAU84075
 ID AAU84075 standard; Peptide; 895 AA.
 XX
 AC AAU84075;
 DT 08-MAY-2002 (first entry)
 XX
 DE S. pneumoniae derived chimeric peptide, VP117.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutain;
 XX BVH-11-2.
 OS Streptococcus pneumoniae.
 XX Synthetic.
 PN W0200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Claim 1; Page -: 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90a
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for

DE *S. pneumoniae* derived chimeric peptide, vp123.

DE **S. pneumoniae derived chimeric peptide, vp123.**

QY 241 SLLGTCNDNTISAEVDSLALLKESQAPIQ-GPQIGOPTLPNNSLATPSPSLPINPCT 299
 Db 236 SLLGTCNDNTISAEVDSLALLKESQAPIQSGPQIGOPTLPNNSLATPSPSLPINPCT 295
 QY 300 SHEKHEEDGYGDANRIIAEDSGFVMSHSDSNHFFFKDLTEBQIKAAQKHLEEVKTS 359
 Db 296 SHEKHEEDGYGDANRIIAEDSGFVMS-----YFFKDLTEBQIKAAQKHLEEVKTS 349
 QY 360 NGLDSLSEHQDYPGNAKEMKDLKIEKIKAGIMQYGVKRESIVVWNEKNAIYPSGD 419
 Db 350 NGLDSLSEHQDYPGNAKEMKDLKIEKIKAGIMQYGVKRESIVVWNEKNAIYPHGD 409
 QY 420 HHADPIDEHKVPVIGCHSHSNTYELFKPEGVAKKEGKNYTTGEELTNVNVNLLKNSTFNNQ 479
 Db 410 HHADPIDEHKVPVIGCHSHSNTYELFKPEGVAKKEGKNYTTGEELTNVNVNLLKNSTFNNQ 469
 QY 480 NPTLANGORVVSFPPELEKLGIMLVKLITPDGKLVLEKYSKVFGEVGNIANFELD 539
 Db 470 NPTLANGORVVSFPPELEKLGIMLVKLITPDGKLVLEKYSKVFGEVGNIANFELD 529
 QY 540 OYPLPGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFA 599
 Db 530 OYPLPGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFA 589
 QY 600 VPKGTDALVRVDEPHGNAYLENNYKVGEIKLPIPKLNGQTTTAGNKIPVTFMANAYLD 659
 Db 590 VPKGTDALVRVDEPHGNAYLENNYKVGEIKLPIPKLNGQTTTAGNKIPVTFMANAYLD 649
 QY 660 NOSTYIVVEPILKEKENTDKPSILPQFKRNKAQENSKLDERVEEPTSEKVEKEKLSG 719
 Db 650 NOSTYIVVEPILKEKENTDKPSILPQFKRNKAQENSKLDERVEEPTSEKVEKEKLSG 709
 QY 720 NSTNSTLEEVPVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVKKNMADF 779
 Db 710 NSTNSTLEEVPVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVKKNMADF 769
 QY 780 TGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPE 839
 Db 770 TGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPE 829
 QY 840 GNVGSDPMLDPALEAPVDPQVEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVIK 899
 Db 830 GNVGSDPMLDPALEAPVDPQVEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVIK 889
 QY 900 NLSDFIA 906
 Db 890 NLSDFIA 896

Search completed: March 27, 2003, 11:54:24
 Job time : 47 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:37 ; Search time 20 Seconds
(without alignments)
1332.858 Million cell updates/sec

Title: US-09-884-465A-332

Perfect score: 4728

Sequence: 1 MQITYTDEIQVAKLAGKYT.....TELRLPSGEVKNLSDFIA 906

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	818.5	17.3	763	4	US-08-961-083-66
2	636	13.5	796	4	US-08-961-083-56
3	431	9.1	447	4	US-08-961-083-182
4	223.5	4.7	10182	4	US-08-134-001C-3159
5	199.5	4.2	1098	4	US-08-923-992A-8
6	195.5	4.1	1128	4	US-08-923-992A-6
7	194.5	4.1	1164	4	US-08-923-992A-2
8	192	4.1	571	4	US-08-961-083-4
9	192	4.1	1164	4	US-08-923-992A-10
10	181.5	4.1	3696	4	US-08-134-001C-5080
11	185.5	3.9	1104	4	US-08-923-992A-4
12	184.5	3.9	1964	2	US-08-790-912-3
13	184.5	3.9	2052	2	US-08-790-912-2
14	182	3.8	984	1	US-08-242-932-2
15	182	3.8	984	1	US-08-714-481-2
16	182	3.8	984	5	PCT-US95-06111-2
17	180.5	3.8	1183	2	US-08-447-031A-2
18	175.5	3.7	1231	3	US-08-904-263A-4
19	175	3.7	1183	4	US-08-134-001C-3530
20	170.5	3.6	1545	4	US-08-286-791-4
21	170.5	3.6	1545	5	PCT-US95-10661A-4
22	170	3.6	1848	4	US-08-236-791-6
23	170	3.6	1848	5	PCT-US95-10661A-6
24	168.5	3.6	3135	1	US-08-323-170B-2
25	168.5	3.6	3135	4	US-08-954-441-2
26	167	3.5	783	6	US-08-168-2
27	165.5	3.5	1181	4	US-09-206-898-23

28 164.5 3.5 962 4 US-09-071-035-246 Sequence 246, App
29 164.5 3.5 962 4 US-09-071-035-250 Sequence 250, App
30 164.5 3.5 962 4 US-09-071-035-254 Sequence 254, App
31 164.5 3.5 962 4 US-09-071-035-470 Sequence 470, App
32 164.5 3.5 962 4 US-09-071-035-474 Sequence 474, App
33 164.5 3.5 962 4 US-09-071-035-478 Sequence 478, App
34 164 3.5 1164 2 US-08-589-756-1 Sequence 1, Appl1
35 164 3.5 1164 4 US-09-206-800-1 Sequence 1, Appl1
36 164 3.5 1164 4 US-09-206-898-1 Sequence 1, Appl1
37 163.5 3.5 1881 4 US-09-233-086-3 Sequence 3, Appl1
38 162 3.4 2123 4 US-08-968-885A-10 Sequence 10, Appl1
39 161 3.4 676 2 US-08-134-001C-4318 Sequence 4318, Ap
40 161 3.4 1167 2 US-08-589-756-2 Sequence 2, Appl1
41 161 3.4 1167 4 US-09-206-800-2 Sequence 2, Appl1
42 161 3.4 1167 4 US-09-206-898-2 Sequence 2, Appl1
43 160.5 3.4 1301 4 US-09-071-035-234 Sequence 234, App
44 160.5 3.4 1301 4 US-09-071-035-238 Sequence 238, App
45 160.5 3.4 1301 4 US-09-071-035-242 Sequence 242, App

ALIGNMENTS

RESULT 1
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 5,349,989
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: FB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-66

Query Match 17.3% Score 818.5; DB 4; Length 763;
Best Local Similarity 66.1% Pred. No. 4.4e-52;
Matches 168; Conservative 16; Mismatches 29; Indels 41; Gaps 5;

Qy 2 QITYTDEIQVAKLAGKYTTEGTFD-----TSWIKKDSLSSEAR 42
|||||
Db 498 QITYTDEIQVAKLAGKYTTEGTFDPRDITSDGDAYTTPHMTSHWIKKDSLSSEAR 557

Qy	2	QIYTTDDEIOVAKLAKYTTEDGYIFD-----TSWIKKDSUSEAER	42
Db	507	QIETDEEVARIAQLAKYTTSDGYIFDEHDIISDEGDAYVTPMHGSHHWIGKDSLSDEK	566
Qy	43	AAAQAYAKERGLTPPSTDHQSGNTAKGAEAIYNRVKAAKVYPLDRMPYNLYQYTVVEKN	102
Db	567	VAAQATYKERGLIPPSPDADVAKNPATGDSAAAIYNRVKEKRIPVLRVLYPMVHEHTVEKN	626

Query Match	4.7%	Score 223.5;	DB 4;	Length 10182;
Best Local Similarity	19.0%;	Fred. No. 1.7e-06;		
Matches 201;	Conservative 169;	Mismatches 409;	Indels 279;	Gaps
QY	37	LSERAAQAQYAKEGLPSTTDHQDSQNTAEKGAEEIYNRVKAAKVPIDRMYPNLQ- 95		
DB	5272	LTAKEDAVASINNLSGLTNEQTKENQAVNQATPDQVANKLRDAE-ALDQSMQTLRD 5329		
QY	96	-----YTVVKNGSLIPSDHYHNKIKPEWDEGL----- 135		
DB	5330	LVNNQNAIHSTSYNFNEDSTQKNTYDINAIDNGSYITG-QHNPELNKSTIDQITSRINTA 5388		
QY	126	-----YEAPKGYSLIEDLATKYVYEPNASHYRNKAQDQSKPDB----- 167		
DB	5389	KNDLHGVEKIQORDQYANQE-IQQLYLNDPQSGEESLVNGSTRSEVSEHLNEAKSLN 5447		
QY	168	---DKHEVSEYPTPE-----SDEKENHAGLNPS---ADNLY----KPSTDTETESEE 211		
DB	5448	NAMKQLRDKVAETNWKQSSDYINDSTEHORGYDQALQEAENIINEIGNPTLNKSEIEQK 5507		
QY	212	AEDTDEABETPGPSTRQNAEMTLGLASILLCTFKDNNTTSAEVDLSALLKESQAPPI 271		
DB	5508	LQQLUTD-----AQNA-----LQSHLELEAKNNAITG-INKLTALNDAORQKAI 5550		
QY	272	QGPOIGOPTLP-----NNSLATPSPSLNPNTGTSHEKEEDY-----G 310		
DB	5551	ENVO-AQOTTIPVNNQULTDRENTAMQALRKVGCQNNVHOOSYFNEDSQPKHNTDNS 5609		
QY	311	FDANRIIAEDSGFVMSHGSDSNHYFFPKDLTBEQIKAAQK-HLEEVKTYSH--NGLDLSL- 366		
DB	5610	VOAGCTTIDKLOPIPNKKNIEQAINQINTOTALSGENKLTDOESTNRQIGESLINT 5669		

APPLICATION NUMBER: US 60/024,707
 FILING DATE: 06-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1098 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-923-992A-8

Query Match 4.29; Score 199.5; DB 4; Length 1098;
 Best Local Similarity 17.68; Pred. No. 3.6e-06;
 Matches 157; Conservative 125; Mismatches 289; Indels 321; Gaps 34;

QY 166 DEDKEHDEVSEPTHP---ESDEKENHAGLNPSADNLYKPSDTTETEEAEADTTDEAIP 222
 DB 12 DSVKTEVAAPKPPSMAOTDGGNSSSELETRMEIPTTDIKKAVEPVKTAGETSAT 71
 QY 223 GTPSIRONAMETLGLKSSL---LLGTRDNNTISAEV-----DSLAL 262
 DB 72 HTGKREKLOQWKNLNDVDNTILSHEQKNEFKTKIDETNDSALLENOFNETNRL 131
 QY 263 -LKESQAPIOGPOIGOPTLPNNSLATSPSLPNTGTSHEKHEEDGVGFANDRIAEDE 321
 DB 132 HIKHEEVEKDKAKAQQTKQSD--TVDLNIDKELNHOKSQEAG-----ITNEDK 182
 QY 322 SGFVMSHGDNNHFFPKDLTE-----BOIK----- 346
 DB 183 DSKAKIEDIRKQAOPDKDAEVKVEELGKLFSSFKAGIDOBIOEHVVKETSEENT 242
 QY 347 -----AAKHLEVKTSNGLDLSLSEHQ-----DYPGNKEMKDLKKIE 387
 DB 243 QKVDEHYANSQNLAKSLEE-----LDKATNEQATQVKNQPLENAKLEIOPLIK 295
 QY 388 EKIAGIKQYGVKRESI-----VVKERNAILYPSGDHHPIDIEHKPVG 433
 DB 296 ETNVLKRAMSSLSQVEKELKHNSANLELDVAKSKEIVREYEGKLNQKNLPKLQLE 355
 QY 434 IGHSHNTELPPEGVAKKGNKYTGEBELTNVNLKNSTFNQNTLANGKRVSEFS 493
 DB 356 -BEAHSKL-----KQVVEDEFKFKTSQVTPKRLKRDLAANE-----NNOOKIELT 402
 QY 494 FPPELEKLGINMLVKLITPDGKLVKSGKVGEGVGNIANFELDQVLPQOTFKYITIA 553
 DB 403 VSPE-----NITVE-----GEDVKFTVT 421
 QY 554 SKDYPEVSDGFTTPTSLAYKMASQTIFFPHAGDTYLRVNPQAPVKGTDALYRVFDE 613
 DB 422 AKSDSKTLD-----FSDLLTKYNPSVS-----DRISTN 450
 QY 614 FPGNAYLENNYKVGKILPIKLNQGTTRT-----AGNKIPVTWANYLDNQSYIVE 667
 DB 451 YKNT---DNHIAEITTKNLNSESQVTLKAKDSDGNVKEPT-----T 493
 QY 668 VPILKENOTDKPSILPOPKRKAQENSKLDEKV-EEPKTSKEVEKELSETGN----- 720
 DB 494 ITVQKEER-----QVPKPEQKDSKTEKVPQEPKSNQNLQELIKSAQOLEKL 545
 QY 721 STSNSTLEVPV-----DPVQEKVAKF-----AESYGMKL 751
 DB 546 EKAIKELAMEQPEIPSNPEYGIQSIWESQKEPIQEAITSFKKIIIGDSSSKYITEHYFNKY 605
 QY 752 ENVLPMNDCTIELYLPBGVVIKKNMADFTGEAPOGENKPKSPENKSVT-----GTVENQ 806
 DB 606 KSHFMNYQLHAQM-----EMLTRKVVQYNNKYPDNAEIKKIFESDMKRTKEDNYGSLND 660

QY 807 PTE-----NK-----PADSLPEAPNEKPVKPNSTDNGLNPEGNVGDPM 849
 DB 661 ALKGYFEKYLTPFNKIKQIVDDFKKVDQDQAPAPI-PENSE-----MD 703
 QY 850 PALEEA-----PAVDPVQOEKLEKFTASYGLGL-----DSVIFNMD 884
 DB 704 QAKKAKIAVSKYMSKVLGDGVHQLKKNHNSKIVDLFKELEAIKQOTIFDID 755

RESULT 6

US-08-923-992A-6
 Sequence 6. Application US/08923992A
 Patent No. 6280738
 GENERAL INFORMATION:
 APPLICANT: Tai, Joseph Y.
 APPLICANT: Blake, Milan S.
 TITLE OF INVENTION: No. 6280738-iga Fc Binding Forms of the Group B
 TITLE OF INVENTION: Streptococcal Beta Antigens
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, N.W., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923,992A
 FILING DATE: 05-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,707
 FILING DATE: 06-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1128 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-923-992A-6

Query Match 4.18; Score 195.5; DB 4; Length 1128;
 Best Local Similarity 17.38; Pred. No. 7.3e-06;
 Matches 154; Conservative 130; Mismatches 296; Indels 309; Gaps 34;

QY 166 DEDKEHDEVSEPTHP---ESDEKENHAGLNPSADNLYKPSDTTETEEAEADTTDEAIP 222
 DB 7 DSVKTEVAAPKPPSMAOTDGGNSSSELETRMEIPTTDIKKAVEPVKTAGETSAT 66
 QY 223 GTPSIRONAMETLGLKSSL---LLGTRDNNTISAEV-----DSLAL 262
 DB 67 DTGKREKLOQWKNLNDVDNTILSHEQKNEFKTKIDETNDSALLENOFNETNRL 126
 QY 263 -LKESQAPIOGPOIGOPTLPN-----NSLATPSPSLPNTGTSHE----- 302
 DB 127 HIKHEEVEKDKAKAQQTKQSDTKVLDNIDKELNHOKSPVEKMAEPKGTITNEDK 186
 QY 303 -----KHEEDGVGFANDRIAEDESGFVMSHGDN-----HYFFKDLTREQ----- 344
 DB 187 LKTIEDINKQAQADKKDAEVKVEELGKLFSSFKAGIDOBIOEHVVKETSEENTQKV 246


```

QY 345 -----IKAAQHLEVKTSNGLDLSLSEHQ-----DYPGNAKEMKOLDKRIEKI 390
Db 247 DEHYANSLONLAQKSLSE-----LDKATNEQATQVKNQFLENQAKLQKMOPLIKETN 299
QY 391 AGINKQYGVKRESI-----VWKENNAIIPSGDHHHADPIDEHPVGLGH 436
Db 300 VKLYKAMESLEQVEKELHNSANLEMDVAKSEIVREYEGKLNQSNLPELKOLE-EE 358
QY 437 SISNYELKPEGVAKKGNVYTGELTNVNLKNSTFNQNFLLANGOKRVSFSPFP 496
Db 359 AHSKL-----KQVEDPRKFKTSQVTPKRVKRDLAANE-----NNQKIELTVSP 406
QY 497 ELEKKLGINMLVKLITPDGKLVKRGVGVGNGVNIANFELDQYLPFGOTFKYTIASKD 556
Db 407 E-----NITVIE-----GDEVKFTVTAKS 425
QY 557 YPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFVAVPKGTDALRVFDEPHG 616
Db 426 DSKTTLID-----FSDLLTKYNPSVS-----DRISTNTKT 454
QY 617 NAYLENNYKVGKILPIPKINOGTIRT-----AGNKIPVTFMANAYLDNOSTYIIVEPI 670
Db 455 NT-----DNHKAIEITKNLKNESQVTLKAKDDSGNVVEKTF-----TIV 497
QY 671 LEKENOTDRPSILPOFRKNAQNSKLDKVK-EEPKTSEKVEKELSETGN-----STS 723
Db 498 QKKEEK-----QVPTPEQKDSKTEKVPQEPKPSNDKNOLELKSAAQOELEKLEKA 549
QY 724 NSTLEVPV-----DPQEKVAKF-----AESYGMKLENV 754
Db 550 IKELMOPPEIPSNPEYGIQKSWESQEKIQAITSFKKIIGDSSKYYTEHYFNKYSKD 609
QY 755 LFNMDGTIELYLPSEGVIKKNMADFTGEAPOGNGENKPSKSVT-----GTVENQPT 809
Db 610 FNNYQLHAQM-----EMLRKVVQYANKYPDNATKIPESDKMRTKEDNYGSLNDALK 664
QY 810 -----NK-----PADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPMDPAL 852
Db 665 GYPERYFLTFPNKIQIVDOLKVKVQDQAPFI-PENSE-----MDQAK 707
QY 853 EEA-----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 884
Db 708 EKAKIAVSKYMSKVLGVGHQHLQKKNHSHKIVDLFKELEAIKQOTIFDID 756

RESULT 7
US-08-923-992A-2
Sequence 2, Application US/08923992A
Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tal, Joseph Y.
APPLICANT: Blake, Milan S.
TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
TITLE OF INVENTION: Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707

```

```

FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1164 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-2

Query Match 4.1%; Score 194.5; DB 4; Length 1164;
Best Local Similarity 19.4%; Pred. No. 9.1e-06;
Matches 166; Conservative 130; Mismatches 316; Indels 245; Gaps 40;

QY 166 DEKEHDEVSEPHTP-----ESDEKENHAGLNPSADNLYKPSDTEETEEAEADTTDEAEP 222
Db 43 DSVKTTTAAKYPSPMAOTDQGNSSSELETTKMBIPTDIKKAVEPVKTAGETSAT 102
QY 223 GTPSIRONAMETTLGLKSSL-----LLGPKDNNTISAQV-----DSLALLKESOPAPIQG 273
Db 103 DTGREQLQOWKNNLANDVNDYILSHEQNEFKTKIDETNDSALLE-----151
QY 274 PQIGOPTLPNNSLATPSPSLPINTPGTSHEHEEDGYGFDANRIAEDESGFVMSHGDSS-- 331
Db 152 -----ENQFNETNRLHLIK---QHEEVEKDKKA-KQOKTKLQSDTKVDLSNIDKEL 198
QY 332 NH-----YFFKDLTEQIKAAQKHLVEKTSNGLDLSLSHEODYPGNAKEMKLDK 384
Db 199 NHQSQVEKMAEQGRTFNEDKDKMLKIEDIRKQAQADKEDAE-----VKVREELGK 252
QY 385 KIEKIAIGIMK--QYGVKRESIVVYKKNKNAIIPSGDHHHADPI-----426
Db 253 LFSSTKAGLDQEIQEHVKET---SSEET---QKVDHYANSLONLAQKSLSELDKATT 306
QY 427 DEHPVGVGHSHSNYELFKPEGVAKKGNKVY--TGEELTNVNLKNSTFNQNFLLA 484
Db 307 NEQATQVKNQFLENQAKLKEIOTLIKETNVLKYKAMESLEQVEKELKHNSANLEDLVA 366
QY 485 NGKRVSPFPPELEKLGKGINMLVKLITPDGKLVKRGVGVGNGVNI-ANFELDQYPL 543
Db 367 KSKEIYR-----EYEGKL--NQSKNL--PELKQLEEEAHSKLGQVDEFRKKFTSQVT 417
QY 544 PQGTFKYTIASKD-----TPE--VSYDG---TFTVPTSLAYKMASQTIFFPHAG 588
Db 418 PKKRVKRDLAANENNQQKIETVSPENITVYEGEDVKFTVTA---KSDSKT---TLDIFS 470
QY 589 DTYLRVNPQFVAVPKGTDALRVFDEPHGNAVLENNYKVGKELPIPKINOGTIRT----- 643
Db 471 DLLTKYNPSVS-----DRISTNYKNT--DNHKAIEITKNLKNESQVTLKAKD 519
QY 644 -AGNKIPVTFMANAYLDNOSTYIIVEPFILEKENOTDKPSILPQFKNKAQNSKLDKRV- 701
Db 520 DSGNVVEKTF-----TIVQKKEK-----QVPTPEQKDSKTEKVP 557
QY 702 EEPKTSKVEKELKSETGN-----STNSNILEEYPTV-----DPVOE 738
Db 558 QEPKSNDRKQLOELIKSAQOELEKLEKAIKELMEQEPSNPEYGIQKSWESQEKI 617
QY 739 KVAKF-----AESYGMKLENVLFNMDGYTIELYLPSEGVIKKNMADFTGEAPOG 786
Db 618 ATTSFKKIIGDSSSKYYTEHYFNKYSKDFMNYQLHAQM-----EMLRKVVQYANKYPDN 672
QY 787 NGENKPSKGVST-----GTVENQPT-----NK-----PADSLPEAPNEK 824
Db 673 AEIKKIFESDMKRTKEDNYGSLNDALKGYFEKYLFTPFNKIKQIVDLDKRVKVEDQAP 732
QY 825 VRPENSTDNGLNPEGNVGSDDPMDPALAEA-----PAVDPVQEKLEKFTASYGLG 875

```

```

203 FSVNKEEVSREIYSTTAPSPRIVEKGGKTKQVIEKQETGVKHKDVQSGAIVPAI-Q 261
QY 541 PYPGOTFKYTIASKDYPEYSDGTFTVPTSLAYKMASOTIFPFGAGTYTLRVNQ--- 597
262 PEP- ---EAVVSDKGEVQP- ---TLPEAVV- ---TDKGET- ---EVQPESPD 300
QY 598 -FAVPGTDLVRYVDFRGNLAYLNNYKVGEEK--LPIPKL-NQGTTFAGNKIPVTFM 653
301 TVVSDKGEPEQVAPLPEYKGN- ---IEOVKPETPVKEKTEGQPEKT- ---EEVPV- --- 347
QY 654 ANAYLDNQSTYIIEVPEFLEKENOTDKPSILPQFKN- ---KAQENSKLDEKVEPEPKTSEK- 709
348 - ---KPEETPVNNEGTEGTSI- ---QEAENPVQAEESTTNSKVV-SPDTSSKN 395
QY 710 - ---VEKEKSETSGNSTNSTSTLEEVTVDPVQVKVAKFAESYGMKLE 752
396 TGEVSSNPDSSTTVSGESNKNPEHDSNKNSEKTVEEVP- ---VNP- --- 437
QY 753 NVLFNMDGTTIELYLPSCGEVIKKMAADFTGAPOGNGENKPSENGKV- ---STGTVENQPT 809
438 - ---NEGTVB- ---GTSNQETEKVPQAEETQTSNGKIANENTGEVSNKPSD 481
QY 810 NKP- ---ADSLPEAPNEKPVKPPENS- ---TDNGMLNPE- ---GNVGSDPMLPALZEAAPVDP 860
482 SKPPVEESNQPEKNGATKPKNSGNTTSSENGQTEPEPSNGSTEDYSTESNTSNGNEE 541
QY 861 VQEKLEFTASYGLGDSVFNMDGTIELRLPSPGVKIKNLSD 903
542 IKQENE- ---LDPKKVEEPEKTYLELR- ---NVSD 568

RESULT 9
US-08-923-992A-10
; Sequence 10, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-1GA FC Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-923-992A-10

Query Match 4.1%; Score 192; DB 4; Length 1164;
Best Local Similarity 17.3%; Pred. No. 1.4e-05;
Matches 148; Conservative 133; Mismatches 283; Indels 298; Gaps 34;

QY 170 EHDVSE-----PHTPSDEKENAGLNPADNLYKFTSTDETEEEAEADTTDAEI 221
Db 42 KHDVKTTEAAKPPYMAQTDOGNSSSELETTTKEIPTTIDIKKAVEPLEKTAGTSA 101

QY 222 PGTPIRONAMETLGLKSSD---LLGTCKNNTISAEV-----DSLAL----- 262
Db 102 TDTGKREKQLOQKNNLKNVHTILSHEKNEFKTIDETNDSALLEENQFNETNRL 161

QY 263 --LKESQAPITQPOIGQPTLPNNSLATPSPSLPINCTSHEKHEEDGYGFDANRIIAD 320
Db 162 LHIQHEVEKDKAKOQKTLKQSD--TKVDSLNDIKELNKHQSKQVETMA--EQLGITNED 218

QY 321 ESGFVMSHGDSNHYFFKDLTEE--QIKA-----AOKHLEEVKTSHG 361
Db 219 KDSMLKTIEDIRKQAOADKEDAEVREBELGLFTSTAGLDQEBQEHVKKETTSEEN 278

QY 362 LQSLSSHEQDYPGN-----AKEMDKLKKIEEKIAGIMQYGVK-----RESIYVKNKN 411
Db 279 TKQVDEH---YFNSLQNLAKSLEELDKATTNEQATQVKNQFLENQAKLKEIQLIKETN 335

QY 412 AITPSPGDHHDADIDEKHPGIGHSHSNYE--LFKPEGVAKKEG----- 455
Db 336 VKLYKA---MSESLQVEKQLKHNQSANLEDLVAKSKEIYREYEGKLNQSKNPELKQL 391

QY 456 -----NKVYTGELTNVNLKNTSPNQNFLLANGOKRYSFSPFPELE 499
Db 392 EBEAKSLKQVVEDRKFKFTSEQTPKRVKRLAANE-----NNQKTELIVSPE-- 443

QY 500 KKLGINMLVKLITPDGVLEKVGKVGEGYGNIANFELDPYLPQGFYTYIAKSKYPE 559
Db 444 -----NITVYE-----GEDLKFTLTAKSDSK 464

QY 560 VSYDTFTVPTSLAYKMASOTIFYPPFHAGDYLYRVNPOFAYPKGTDALVRVDFEFGHAY 619
Db 465 TFLD-----FSDLITKYNPSVS-----DRISTNKTNT- 492

QY 620 LENNYKVGKILPIKPLNOGTTT-----AGNKIPVTFMANAYLDNSTYIVEPILEK 673
Db 493 --DNHKAIEITKLNKESQITVLKARDSGNVVKYF-----TITVOKK 536

QY 674 ENOTKPSILPOFKRNKAQENKSLDEKV--EEPKTSKEVEKEKLSGTGN-----STSNST 726
Db 537 EEK-----QVPKTPQKDSKEEKVPQEPKSNQKQLOELIKSAOQOLEKLEKAIKE 588

QY 727 LEEVPTV-----DPVQEKVAKF-----AESYGMKLENVLFN 757
Db 589 LMEQPIPSNPEYKQKTSWESQKEPIQBAITSFKKIIGDSSSKYTYTEHYFNKYKSDFMN 648

QY 758 MDGTIELYLPSEVIAKKNMADFTGEAPOGNGENKSPENGKST-----GTVENQPTPE--- 809
Db 649 YOLHAQM-----EMLTRKVQVQINKYPDQNAEIKKIPESDMKRTKEDNYGSLENDALKGYF 703

QY 810 -----NK-----PADSLAPNEKVPKPNSTDNGLNPEGVNVDPMLDPALBEA 855
Db 704 EKYFLTFPNKIKQIVDDLDKQVEQDQAPFI--PENSE-----MDQAKERA 746

QY 856 -----PAVDPVQEKLEK 867
Db 747 KIAVSKYMSKVLGVQVHQLQK 767

RESULT 10
US-09-134-001C-5080
; Sequence 5080. Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-stamm et al

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: GTC-007
;; CURRENT APPLICATION NUMBER: US/09/134,001C
;; CURRENT FILING DATE: 1998-08-13
;; PRIOR APPLICATION NUMBER: US 60/064,964
;; PRIOR FILING DATE: 1997-11-08
;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 5080
;; LENGTH: 3696
;; TYPE: PRT
;; ORGANISM: Staphylococcus epidermidis
;; US-09-134-001C-5080

Query Match 4.1%; Score 191.5; DB 4; Length 3696;
Best Local Similarity 18.9%; Pred. No. 8.4e-05;
Matches 180; Conservative 143; Mismatches 344; Indels 287; Gaps 41;

QY 6 TDEIOVAK--LAGKYTTEDGYLFTDTSWIKKDSLSAERAAQAAYAKERGLTPPSTDHOD 63
Db 2041 TDEKEAAKQLVTKLINEQIKIHEST---QD--NOVDNVKAQAITAIKLINANAKHQD 2095

QY 64 SGNTEARGAEIYRNRYKAARKVPLDRMPYNLOTVYKNGSLIIPSYDHYHNKIFWFDE 123
Db 2096 AINILTNLAESKSDIRA-----NQDATTEKNTAI----- 2126

QY 124 GLYEAPKGYGLELLATVYVYVPEPNASDHVRKNKADQDSKPD----- 167
Db 2127 -----QSIDDTLAQARNNINGANTNALVDENLEGGKQLQRIVLSQTQKQAKADIA 2178

QY 168 ---DKHEDEYSETPHESDEKENHAGLNPADNLYKFTSTDETEEEA---EDTTDEAEI 221
Db 2179 QATGQOQSTTDQNONATTEKQE-----ALERNGETNGVNDRIQAALANQVTD----- 2229

QY 222 PGTPIRONAMETLGLKSSLLGTDKNTTISAEVDSLLALLKESQAPITQGGPQIGQPTL 281
Db 2230 -----KNNILETIRNVEPIVIRKANEIIRKKAARQTTLINQODATLEEKQIALGKL 2283

QY 282 ---PNNSLATPSPS-----LPINPGTS--HEKHEDEGYGFDANRIIAEDSGFVMSHG 329
Db 2284 EEVKNALNQVSAHSNNDVKNIAENNGIAKISEVHPETIIRKNAKQIEQDQASQIDTIN 2343

QY 330 DSNHYFFKDLTEBQIKAAQK---HLEEVKTSHGSLDLSHSHEDQYYPGNKEMKDLKK 385
Db 2344 ANN-----KSTNEEKSAAIDRVNVAKIDAINNITNATTTQLVNDAKNSGNT----- 2389

QY 386 IEEKIAGIMQYGVKRESIVV---NKEKNALYPSGDHHDADIDEKHPVGIHSHSNY 441
Db 2390 ---SISQILPSTAVKTNALAAALASEAKNNAIL-----DQTPNATAEEK-----EANN 2435

QY 442 ELFKPEGVAKKEGN--KVYTGELTNV-----VNLKNTSPNNQNTLANGQ 487
Db 2436 KYDRLQE---EADANILKAHTTDEVNINKNQAVQVNAQVVEIKQNVKNQNLQFIDNQ 2492

QY 488 KRVSEFSP-PELEKLGINMLVKLITPDGVLEKVGKVGEGYGNI--ANFELDQPLPG 545
Db 2493 KKIIENTPDATLEEKAEANRLQ-----NYLTSTS-----DEIANVDHNEVDQ----- 2536

QY 546 QTFKTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDYLYRVNPOFAYPKGT 605
Db 2537 -----ALDKARPKIE-----EIVPO--VSKKRD 2557

QY 606 ALVRVDFEFGHAYLNNYKYGEIKLPIKPLNOGTTTGTAGNKIPVTFMANAYLQNO--STY 664
Db 2558 VLNAIQEAF-----NSQTOEIQ-----ENQEATN 2581

QY 665 IVEVPILEKQNTDPSILPQFKN--KAQNSKLD-----EKVEPKTSKVEK 712
Db 2582 EEKTEALAKING-----LLNQAKVNDIQAKNSKQVDSAKTSIQDIEQIHPOTKATGR 2636

QY 713 EKLSETGNSTNSTLEEVP--TVDPVQEKVAKFAE-----SYGMKLENVLFNM 758

Db 2637 HRLNEKAP--QOSTIATPNSTIEERQASAKLOEVLKAKAIKIDKQTNDDVKTVMNG 2695
QY 759 DGTIELPSPSEVKKNNKADTGAPQNGENKPSKGVSTGVNPTENK--PADSL 816
Db 2696 IAEIENILPATVTKADKADYNAEKEQN-----LQINSNDEATTEKLVASDNL 2745
QY 817 PEAPNEKPVKPNSTNGKMLNPNVNGSDPMLD--PALEAP-ADVDPQEKLEK 867
Db 2746 NHVYETTNOAIEDAPDTNQNVNVEKNKGITGIRDIOPLVVKPTAKSKIESAVEK 2799

RESULT 11
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-4

Query Match 3.9%; Score 185.5; DB 4; Length 1104;
Best Local Similarity 18.9%; Pred. No. 3.9e-05;
Matches 162; Conservative 131; Mismatches 319; Indels 245; Gaps 39;

QY 166 DEKEHDEVEPTHP---ESDERENHAGLNPASDNLKPKSTDETEEEAEDTDEAEP 222
Db 12 DSVKTEVAAPKPPSMAQDQGNSSSELETETRMIPTHIKKAVEPEVKTAGETSAT 71
QY 223 GTPSIRONAMETITGLKSSL---LLGPKDNNITSAEY-----DSLALLKESOPAPIQG 273
Db 72 DTGRRKQLOQWNNKNDVNTILSHEQNEFKTKIDETNDSALLEL----- 120
QY 274 PQIGQPTLPNLSATPSPSPINPGTSHKEHEGYGFDANRIAEDESGFVMSHGD-- 331
Db 121 -----ENQFNETNRLLRHQRHEVEKHKRP---NQKTLKQSDTKVDSLNDIKEL 167
QY 332 NYHFFKDD-----LTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKEMKOLDK 384

Db 168 NHQSOVEAMAEQAGITNEDKDSMLKTIEDIRKQAQADKKEDAE-----VKVREELGK 221
QY 385 KIEKLIAGIMQY--GVRESIVVKNKENAIYPSGDHHDHAPD----- 426
Db 222 LFSSTKAGLQOOIOEHVAKET---SSEENT---QKVDSHYANSLQNTAKQSLSEELDKATT 275
QY 427 DEHPVPGIGHSHSNYELFKPEEGVAKKGNKVY--TGEEELNVNVLKNSTFNQNFLLA 484
Db 276 NEQATQVKNQFLENAQKLKEIPLIKETNVLKRAMSSELEQVEKELAHSEANLQDIYA 335
QY 485 NQQRVSFSPFPELEKLGIMLVKLITPDQVKLEKVGKVGEGVGN1-ANFELDQPYL 543
Db 336 KSKEIVR-----EYEGKL--NQSKNL--PELKQLEEEAHSKLQVVEFRKFKFTSEQVT 386
QY 544 PGOTFKYTIASKD-----YPE--VSYDG--TPTVTSLAYKMASOTIYPPFHAG 588
Db 387 PKRVRKRDLAANENNOQKIETVSPENITYVEGEDVKTFTA---KSDSKT---TLDFS 439
QY 589 DTYLRVNPQFAVPRGTDAVRFVDFHGNAYLENNYKVGKIKLPKPKNOGTTT----- 643
Db 440 DLLTKYNPSV-----DRASTNYKTNT---DNHKAETITKLNKLNQSTVTLKAKD 488
QY 644 -AGNKIPVTMANAYLQNGSTYIIVEVPILEKENOTDKPSILPQPKRNKAQNSKLDKRV- 701
Db 489 DSGNVVEKTF-----TITVQKKEK-----QVPKTPQKHKSKEQNV 526
QY 702 EEPKTSKVEKEKLETCGN-----STSNSTLEEVPTV-----DPVQE 738
Db 527 QEPKSDKNQLOELIKSAQOELEKLEKAIKELMEQEPISNPEYGIQKSIWESKEPIQE 586
QY 739 KVAKF-----AESYGMLENVFNNDGTIELYPSGVEIKNNMADTFGEAPOG 786
Db 587 AITSFNKIIGDSSSKSYTEHYFNKSHFNVOLHAQN-----EILFKVQVYNNKYPDN 641
QY 787 NGENKPSNGKVS-----GIVENQPT-----NK-----PADSLPEAPNEKP 824
Db 642 AEIKKIFESDMKRTKEDNYGSLNDALKGPERYFLTPFNKIKOIVDDLDKVKVQDQAP 701
QY 825 VKPENTDNGMLNPEGNVGSDPMDLPALEA-----PAVDPOVEKLEKFTASYGLG 875
Db 702 I-PENSE-----MDQAKAKIAVSKYMSKVLGVDGVHQLKKNHSKIVD 744
QY 876 L-----DSVIFNMD 884
Db 745 LKLEAIKQOTIFDID 761

RESULT 12
US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997

CLASSIFICATION: 424
PRIOR APPLICATION DATA: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 3.98; Score 184.5; DB 2; Length 1964;
Best Local Similarity 21.38; Pred. No. 0.00011;
Matches 165; Conservative 103; Mismatches 269; Indels 237; Gaps 41;

QY 156 KNAQDDSKPDDEKE--HDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDEETEERAE 213
DB 51 KYADSELSESEKQGVYD---IPTVENDDETYL-----VYKLSNQ----- 92

QY 214 DTTDAEPTGPSIRQNAEMTLGLKSLLLGTRDNNTISAEVDSLALLKESQAPATQ 273
DB 93 ----LAELPNTGS--KNERQALVAGASLAALGILIFAVSKKKVKNKTVLHLVLVAGMGNG 146

QY 274 POIGOPTLPNLSLATPSPSLPINQTSHEK-----HEEDGYGFDAENRIIAEDSGFVMSHG 329
DB 147 VLVSVALENHLLNNTDYEL---TSGEKPLPKEISGYTYIGYIKGKRTSDPEVSN- 202

QY 330 DSNHYFFKKDLTEEQIKAAQKHEEVKTSNGLDLSLSEHQDYPGNNAKEMKDLKKIEEK 389
DB 203 -----QEKSAATPKQKQVDYN---VTFNFDHPSTVQAIQEQTPVSTK 244

QY 390 TAGIMQYGVKRESIVVKNENAIYPSGDHHDADIDEHKPVGIGHSHSNYELFKPEG 449
DB 245 PTEV-----QVVEKPFSTELINPRKEKQSSDSQOLA-----EKKNET-KREEK 289

QY 450 VAKKEGKVTY-----GELTNVNV-----LLKNSTN-----NONTLANGOKRVSF 492
DB 290 ISPKRTGVTINPQDEVLSQLNKPELLYREETIETKIDFQEEIQENPDLABGTVRV-- 347

QY 493 SPPELEKKLAGINM-LVKLIIT-----PDGKVLKVSQK--VFGE----- 528
DB 348 ---KQEGKLGKVEIVAFISVKNKEVSREIVSTSTTAPSPRIVEKGYKQVKEQPET 403

QY 529 GV-----GNITANFELDQYLPQGTFFYTIASKDYPEVSDGTFTVPTSLAYKMASOTI 581
DB 404 GVEHRDVQSGAIVEPAI-QPELP-----EAVVSDKGEVQVOP-----TLPEAVV----- 446

QY 582 FYPPHAGDTYLRVNPQ-----FAVPKGTDALVRVDFEFGNAYLENNYKVGELK--LPK 635
DB 447 ----TKGET--EVQPESTVTVSDKGEPEQVAPLPEYKGN-----IEQVKPTPEVK 493

QY 636 L-NQGTETAGNKIPVTFMANAYLDNQSTIYIVFVPILEKENOTDKPSILPQKRN---KA 691
DB 494 TREOQPEKT--EEVPV-----KPTETVPNNEGTEGTSI--QEAENPVQPA 537

QY 692 QNSKLDKVEPKTSSEKVK-----EKEKLSGTGNSNSTLEEVPTVD 734
DB 538 EESTNSEKVPDTSSENTGEVSSNPSTTSVGSNKPENHNSKNENSEKTVVEVP-VN 596

QY 735 PVQEKVAKFAESYGMKLENVLFWNDGTIELPLSPSEVIVKKNWADTGEAPQNGENKSE 794
DB 597 P-----NEGTVB-----GTSNOETEKVPQPAEETQTN 623

QY 795 NGKV---STGTVENQPTENKP--ADSLPEAPNEKPKVPKPS-----TDNGMLNPE 839
DB 624 SGKIANETGEVSNKPSDSKPPVEESNQPKNGTATKPSNGNTTSNGQTEPE 677

RESULT 13
US-08-790-912-2
Sequence 2, Application US/08790912
Patent No. 5976542
GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
APPLICANT: Plaut, Andrew G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08790,912
APPLICATION NUMBER: US/08790,912
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-2

Query Match 3.98; Score 184.5; DB 2; Length 2052;
Best Local Similarity 21.38; Pred. No. 0.00012;
Matches 165; Conservative 103; Mismatches 269; Indels 237; Gaps 41;

QY 156 KNAQDDSKPDDEKE--HDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDEETEERAE 213
DB 124 KYADSELSESEKQGVYD---IPTVENDDETYL-----VYKLSNQ----- 165

QY 214 DTTDAEPTGPSIRQNAEMTLGLKSLLLGTRDNNTISAEVDSLALLKESQAPATQ 273
DB 166 ----LAELPNTGS--KNERQALVAGASLAALGILIFAVSKKKVKNKTVLHLVLVAGMGNG 219

QY 274 POIGOPTLPNLSLATPSPSLPINQTSHEK-----HEEDGYGFDAENRIIAEDSGFVMSHG 329
DB 220 VLVSVALENHLLNNTDYEL---TSGEKPLPKEISGYTYIGYIKGKRTSDPEVSN- 275

QY 330 DSNHYFFKKDLTEEQIKAAQKHEEVKTSNGLDLSLSEHQDYPGNNAKEMKDLKKIEEK 389
DB 276 -----QEKSAATPKQKQVDYN---VTFNFDHPSTVQAIQEQTPVSTK 317

QY 390 TAGIMQYGVKRESIVVKNENAIYPSGDHHDADIDEHKPVGIGHSHSNYELFKPEG 449
DB 318 PTEV-----QVVEKPFSTELINPRKEKQSSDSQOLA-----EKKNET-KREEK 362

QY 450 VAKBEGNKVYT-----GEELTNVNV-----LLKNSTFN-----NONFTLANGKRVSE 492
 DB 363 ISPKKCTGNTLNPQDEVLGOLANKPELLYREETIETRIDFQEEIQENPDIAEGTVRV-- 420
 QY 493 SPPPLEKLGKLN--LVKLI-----PDGKVLKYSVK--VFGE----- 528
 DB 421 -----KOEGLKGVKIVRIFSVNKEEVSREIVSVSTTAPSPRIVEKGGTKKQVKEQPET 476
 QY 529 GV-----GNIANFELDQYLPQGTFFYTIASKDYPEVSYDGTFTVPTSLAYKASQTI 581
 DB 477 GVEHDKVQGAIVEPAI--QPELP-----EAVSDKGEPEVQP-----TLPEAVV----- 519
 QY 582 FYPFHAGDTYLRVNPQ-----FAVPGTALYRVFDEFNGNAYLENNYKVGK--LPIPK 635
 DB 520 -----TDKGET--EVOPESDPTVYSDKGEPEQVAPLPEYKGN-----IEQVKTPTPEK 566
 QY 636 L--NOGTVTRTAGNKIPVTFMANAYLDNQSYIIVEPFILEKENQTDKPSILPOFKRN---KA 691
 DB 567 TEOQPEKT--EEVPV-----KPIETVPVNEGTEGTSI--QEAENPVQPA 610
 QY 692 QNSKLDKVEBPKTSEKV-----EKEKLSGTGNSNSTLEBVPVTD 734
 DB 611 EESTNSEKVPDTSSENGEVSNNPDSSTTSVGSSEKPNPEHNSKENSEKTVEEVP--VN 669
 QY 735 PVQEKVAKFAESYGMKLENVLFMDGTIELYLPSCGEVIKKNMADFTGAPOGNGENKPS 794
 DB 670 P-----NEGTVB-----GTSNETEKPQVPAEETQTN 696
 QY 795 NGKV-----STGTVENOPTENK--ADSLPEAPNEKPKVPKENS-----TDNGMLNPE 839
 DB 697 SGKIANETGEVSNRPSDSKPPVEESNQPEKNGTKPKPENSNGTTSENGQTEPE 750

RESULT 14

US-08-242-932-2
 ; Sequence 2, Application US/08242932
 ; Patent No. 5595740
 ; GENERAL INFORMATION:
 ; APPLICANT: Brady, L. Jeannine
 ; TITLE OF INVENTION: Cloning of No. 5595740-Iga Fc Binding Forms of
 ; the Group B Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Saliwanchik & Saliwanchik
 ; STREET: 2421 N.W. 41st Street, Suite A-1
 ; CITY: Gainesville
 ; STATE: FL
 ; COUNTRY: USA
 ; ZIP: 32606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/242,932
 ; FILING DATE: 16-MAY-1994
 ; CLASSIFICATION: A35
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Saliwanchik, David R.
 ; REGISTRATION NUMBER: 31,794
 ; REFERENCE/DOCKET NUMBER: UFI42
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 904-375-8100
 ; TELEFAX: 904-372-5800
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 984 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ;

Query Match 3.88; Score 182; DB 1; Length 984;
 Best Local Similarity 17.2%; Pred. No. 5.9e-05;
 Matches 141; Conservative 117; Mismatches 247; Indels 316; Gaps 33;
 QY 166 DEKEHDEVSEPTHP---ESDEKENHAGLNPSADNLYKPSSTDTETEETEEAEADTTDAEIP 222
 DB 6 DDSVKITEVAAKYPSPMAQTDQGNSSSSSELTTKMEIPTDIDIKKAVEPVKTAGETS-- 63
 QY 223 GTPSIRONAMETITGLKSSLLGLTKONTISAEVDSLALLKESQAPIGQPIGQITLP 282
 DB 64 -----ATDTGKREKLOQMKNN--LKNDVDNTI----- 89
 QY 283 NNSLATPSPLINPGTSHKHEEDGYGFDANRIIAEDSGFYMSHGSDSNHYFFKKDLTE 342
 DB 90 -----LSHEQKNE-----FKTKIDE 104
 QY 343 EQTKAAQKHLEEVKTSNGLDLSSSHEQ--DYPGNAKEMKOL-----DKKIEKIAIGIMKQYG 398
 DB 105 TNSDALLELENOFNETNRLLIHIKQHEEVEKKAQKQTKLQSDTKVD--LSNIDKELN 162
 QY 399 VKRESI-----VVKENKNAIIPSGD--HHADPTDEHKPVGI-----GHSHSNYELF 444
 DB 163 HOKSQVEKMAEQGITNEDKDSMLKTIEDIRKQAOQADKKEDAEVKQLEEEAHSL--- 219
 QY 445 KPEEGVAKKEGKNTYTGELTNVYNLLKNSTFNQNFLLANGOKRVSPFPPELEKKLGI 504
 DB 220 ---KQVVEDFRKKFKTSEQVTPKKRVKRDLAANE-----NNOQKIELTVSPE----- 263
 QY 505 NMLVKLITPDGKVLKSVKGVGEGVGNIANFELDQYLPQGTFFYTIASKDYPEVSYDG 564
 DB 264 -----NITYE-----GEDYKFTVTAKSDSKITLD- 288
 QY 565 TFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAPKPGTDALVRVDEFEHGNAYLENNY 624
 DB 289 -----FSDLITKYNPSV-----DRISTNYKNT---DNH 315
 QY 625 KVEIKLPIPKLNQGTTRT-----AGNKIPVTFMANAYLDNOSTYIVVEPILEKENQTD 678
 DB 316 KIAETIKNLKNSQTVTLKAKDDSGNVVEKTF-----TITVQKKEK-- 359
 QY 679 KPSILPQKRNKAQENSKLDEK--EETKTSKVEKEKLSGTGN-----STSNSTLEVP 731
 DB 360 -----QVPKTPQKDSKTEEKVPEPKNSDKNLQLELIKSAQOQLEKLEKAIKELMEQP 413
 QY 732 TV-----DPVQEKVAKF-----AESYGMKLENVLFNMDGTI 762
 DB 414 EIPSNPEYGIKSIWESQKEPIQEAITSFKKLIIGDSSSKYITEHYFNKYKSDFNMYQLHA 473
 QY 763 ELYLPSGEVIKKNMADFTGEAPQNGENKPSNGKVS-----GTVENQPT----- 809
 DB 474 OM-----EMLTRKVVQYMNKYPDIAEIKKIFESDMKRTKEDNYGSLNDALKGYFEKYL 528
 QY 810 ---NK-----PADSLPEAPNEKPKVPKPSNDNGMLNPEGNVSGDPMPLDPALEEA----- 855
 DB 529 TPFNKIKQIVDDLDKKVQDQAPPI-PENSE-----MDQAKERAKIAVS 571
 QY 856 ----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 884
 DB 572 KYMSKVLGQVHQHLQKKNKSKIVDLFKELEAIKQOTIFDID 612
 RESULT 15
 US-08-714-481-2
 ; Sequence 2, Application US/08714481
 ; Patent No. 5766606
 ; GENERAL INFORMATION:
 ; APPLICANT: Brady, L. Jeannine
 ; TITLE OF INVENTION: Cloning of No. 5766606-Iga Fc Binding Forms of
 ; the Group B Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Saliwanchik & Saliwanchik

QY 679 KPSILPQFRRNKAQENSKLDKRV-EEPKTSKEKVEKLSGTGN-----STSNSTLEEV 731